

ATTN: Ed. Hart.

CFFE

Access DB# 103454

RECEIVED

## SEARCH REQUEST FORM

SEP -9 2003

Scientific and Technical Information Center

STIC/BIOTECH DIVISION  
(STIC)

Requester's Full Name: Celine Qian Examiner #: 78710 Date: 9/8/03  
 Art Unit: 1636 Phone Number 306-0283 Serial Number: 091877935  
 Mail Box and Bldg/Room Location: 11E012/11C10 Results Format Preferred (circle): PAPER DISK E-MAIL

If more than one search is submitted, please prioritize searches in order of need.

\*\*\*\*\*

Please provide a detailed statement of the search topic, and describe as specifically as possible the subject matter to be searched. Include the elected species or structures, keywords, synonyms, acronyms, and registry numbers, and combine with the concept or utility of the invention. Define any terms that may have a special meaning. Give examples or relevant citations, authors, etc, if known. Please attach a copy of the cover sheet, pertinent claims, and abstract.

Title of Invention: Regulatory seq of mouse villin gene

Inventors (please provide full names): D. Pinto et al.

Earliest Priority Filing Date: 6/8/2001

\*For Sequence Searches Only\* Please include all pertinent information (parent, child, divisional, or issued patent numbers) along with the appropriate serial number.

*Please search for SEQ ID NO: 1 & oligo search.*

Edward Hart  
 Technical Info. Specialist  
 STIC/Biotech  
 CMI 6B02 Tel: 305-9203

## STAFF USE ONLY

Searcher: \_\_\_\_\_  
 Searcher Phone #: \_\_\_\_\_  
 Searcher Location: \_\_\_\_\_  
 Date Searcher Picked Up: 9/12/03  
 Date Completed: 9/15/03  
 Searcher Prep & Review Time: \_\_\_\_\_  
 Clerical Prep Time: \_\_\_\_\_  
 Online Time: \_\_\_\_\_

## Type of Search

NA Sequence (#) 2  
 AA Sequence (#) \_\_\_\_\_  
 Structure (#) \_\_\_\_\_  
 Bibliographic \_\_\_\_\_  
 Litigation \_\_\_\_\_  
 Fulltext \_\_\_\_\_  
 Patent Family \_\_\_\_\_  
 Other \_\_\_\_\_

## Vendors and cost where applicable

STN \_\_\_\_\_  
 Dialog \_\_\_\_\_  
 Questel/Orbit \_\_\_\_\_  
 Dr.Link \_\_\_\_\_  
 Lexis/Nexis \_\_\_\_\_  
 Sequence Systems 04/03  
 WWW/Internet \_\_\_\_\_  
 Other (specify) \_\_\_\_\_

**THIS PAGE BLANK (USPTO)**

Pending Nucleic Acid and/or Pending Amino Acid database searches now generate two sets of results. These databases were split into two parts to reduce the time needed to update the databases daily. The split freed up more machine time for processing searches.

Searches run against the Nucleic Acid Pending database produce two sets of results, with the extensions, **.rnpm** and **.rnpn**.

Searches run against the Amino Acid Pending database produce two sets of results, with the extensions, **.rapm** and **.rapn**.

*The Pending database search results should not be left in the case because they contain data that is confidential.*

**THIS PAGE BLANK (USPTO)**





# **STIC Search Report**

## **Biotech-Chem Library**

**STIC Database Tracking Number: 103454**

**TO: Celine Qian**  
**Location: CM-1/11C01/11E12**  
**Art Unit: 1636**  
**Monday, September 15, 2003**

**Case Serial Number: 09/877935**

**From: Edward Hart**  
**Location: Biotech-Chem Library**  
**CM1-6B02**  
**Phone: 305-9203**

**edward.hart@uspto.gov**

### **Search Notes**

Examiner Qian,

Here are the results of the search you requested.

Please feel free to contact me if you have any questions.

Edward Hart

**THIS PAGE BLANK (USPTO)**



# STIC SEARCH RESULTS

## Biotech-Chem Library

Questions about the scope or the results of the search? Contact *the searcher or contact:*

Mary Hale, Information Branch Supervisor  
308-4258, CM1-1E01

## Voluntary Results Feedback Form

➤ I am an examiner in Workgroup:  Example: 1610

➤ Relevant prior art **found**, search results used as follows:

- ☐ 102 rejection
- ☐ 103 rejection
- ☐ Cited as being of interest.
- ☐ Helped examiner better understand the invention.
- ☐ Helped examiner better understand the state of the art in their technology.

Types of relevant prior art found:

- ☐ Foreign Patent(s)
- ☐ Non-Patent Literature  
(journal articles, conference proceedings, new product announcements etc.)

➤ Relevant prior art **not found**:

- ☐ Results verified the lack of relevant prior art (helped determine patentability).
- ☐ Results were not useful in determining patentability or understanding the invention.

Comments:

Drop off or send completed forms to STIC/Biotech-Chem Library CM1 - Circ. Desk



**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 13, 2003, 19:07:58 : Search time 16404 Seconds  
(without alignments)  
13327.162 Million cell updates/sec

Title: US-09-877-935-1

Perfect score: 8995

Sequence: 1 gatctgtgcacacagagaca.....ctctagctctgcacacatg 8995

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estlin:\*  
4: em\_estmu:\*  
5: em\_estrov:\*  
6: em\_estropl:\*  
7: em\_estro:\*  
8: em\_hic:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_hic:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: em\_gss\_hum:\*  
18: em\_gss\_inv:\*  
19: em\_gss\_pln:\*  
20: em\_gss\_vrt:\*  
21: em\_gss\_fun:\*  
22: em\_gss\_mam:\*  
23: em\_gss\_mus:\*  
24: em\_gss\_pro:\*  
25: em\_gss\_rtd:\*  
26: em\_gss\_phg:\*  
27: em\_gss\_vrl:\*  
28: gb\_gss1:\*  
29: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
C 1	498	5.5	507	28	AZ354198	AZ354198 1M0093E23
C 2	445.6	5.0	495	28	BH086621	BH086621 RPT-24-2
C 3	322.6	3.6	336	28	A2657211	A2657211 1M0533615
C 4	291.8	3.2	307	28	AZ354559	AZ354559 1M0093E23

C 5	280	3.1	291	28	AZ821333	AZ821333 2M0094K06
C 6	257.2	2.9	639	28	AZ821403	AZ821403 2M0094I08
C 7	176.4	2.0	641	10	BB612611	BB612611 B0112A03-
C 8	175.8	2.0	589	14	CA884544	CA884544 B0112A03-
C 9	175.8	2.0	4682	11	AK052834	AK052834 Mus muscu
C 10	174	1.9	423	14	CB178219	CB178219 In95B04.Y
C 11	173.4	1.9	399	12	B1465446	B1465446 JeltC03.Y
C 12	172.4	1.9	383	12	BM875096	BM875096 JeltC03.Y
C 13	170.4	1.9	747	9	AV307409	AV307409 AV307409
C 14	166.8	1.9	336	13	BY324393	BY324393 BY324393
C 15	166.8	1.9	346	13	BY327420	BY327420 BY327420
C 16	166.4	1.8	449	9	A1836762	A1836762 UI-M-AJO-
C 17	166.4	1.8	3090	11	AK037096	AK037096 Mus muscu
C 18	165.4	1.8	703	28	BH108871	BH108871 RPT-24-3
C 19	165.2	1.8	510	28	A2406837	A2406837 1M0176C14
C 20	165.2	1.8	1491	11	BC024585	BC024585 Mus muscu
C 21	164.6	1.8	682	28	A2313191	A2313191 1M0059M18
C 22	162.6	1.8	502	14	CA947391	CA947391 In82d03.Y
C 23	161	1.8	623	28	A2766497	A2766497 1M0564D10
C 24	161	1.8	626	28	A2369208	A2369208 1M0119K10
C 25	161	1.8	680	28	A2964931	A2964931 2M0234L18
C 26	160.6	1.8	634	28	A2579634	A2579634 1M0367B16
C 27	160	1.8	558	13	BQ478801	BQ478801 1K61C05.Y
C 28	159.2	1.8	479	9	AA238960	AA238960 mx9410.r
C 29	159.2	1.8	666	10	BG670252	BG670252 DRNAC06
C 30	159	1.8	538	28	A2699259	A2699259 RPT-23-2
C 31	158.4	1.8	630	10	BB655081	BB655081 BB655081
C 32	158.4	1.8	782	28	BH069206	BH069206 RPT-24-2
C 33	157.8	1.8	837	10	BB620553	BB620553 BB620553
C 34	157.8	1.8	697	11	AK077869	AK077869 Mus muscu
C 35	157.6	1.8	409	28	A2075194	A2075194 RPT-23-4
C 36	157.6	1.8	504	28	A2980279	A2980279 2M0257G14
C 37	157.4	1.7	518	14	CA879083	CA879083 K0917F12
C 38	157.4	1.7	518	28	A2034666	A2034666 RPT-23-2
C 39	157.2	1.7	590	28	A2645050	A2645050 1M0510G18
C 40	157.2	1.7	3874	11	AK083490	AK083490 Mus muscu
C 41	156.4	1.7	607	14	BY721781	BY721781 BB673144
C 42	156.2	1.7	424	10	BB673144	BB673144 BB673144
C 43	156.2	1.7	745	12	BM933152	BM933152 UI-M-BH3-
C 44	156.2	1.7	2114	11	AK008714	AK008714 Mus muscu
C 45	156	1.7	627	28	A2809859	A2809859 2M0074K03

#### ALIGNMENTS

RESULT 1  
LOCUS AZ354198 507 bp DNA linear GSS 02-OCT-2000  
DEFINITION 1M0093E23F Mouse 10kb plasmid UUGC1M library Mus musculus genomic  
clone UUGC1M0093E23 F, genomic survey sequence.

ACCESSION AZ354198  
VERSION AZ354198.1 GI:10465402

KEYWORDS GSS.  
SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
1 (bases 1 to 507)

AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamill,C.,  
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Rilly,  
M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.  
and Wright,D., Weiss,R.

TITLE Mouse whole genome scaffolding with paired end reads from 10kb  
plasmid inserts

JOURNAL Unpublished  
COMMENT Contact: Robert B. Weiss  
University of Utah Genome Center

University of Utah  
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT  
84112, USA  
Tel: 801 585 5606  
Fax: 801 585 7177

Email: ddunn@genetics.utah.edu  
Insert Length: 10000 Sca Error: 0.00  
Plate: 0093 row: E column: 23  
Seq primer: CGTTGTAAACGACGCGCCACT  
Class: plasmid ends  
High quality sequence stop: 507.

## FEATURES

## SOURCE

Location/Qualifiers

1..507

/organism="Mus musculus"

/mol\_type="genomic DNA"

/strain="C57BL/6J"

/db\_xref="taxon:10090"

/clone="UGC1M0093E23"

/sex="Male"

/lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"

/clone\_1lb="Mouse 10kb plasmid UGC1M library"

/note="Vector: PMD42nv. Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of PMD42 (g1473214[bp]AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 125 a 139 c 105 g 138 t  
ORIGIN

Query Match 5.5%; Score 498; DB 28; Length 507;  
Best Local Similarity 100.0%; Pred. No. 1.4e-43;

Matches 498; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

8431 CATCCAGAGTGGCTAGCCAGGAGACTCAGGATATGCTGCTCTCCCTCCACAGTCT 8490

507 CATCCAGAGTGGCTAGCCAGGAGACTCAGGATATGCTGCTCTCCCTCCACAGTCT 448

8491 AGAATTACAGGACATACATCTCTGGAATTTTAACTGAACTCTGAGATAGAGCA 8550

447 AGAATTACAGGACATACATCTCTGGAATTTTAACTGAACTCTGAGATAGAGCA 388

8551 GGCACCTCTACCAATGAGAGGCTCTTTTGTGTTGGTTTGTCTCTGATTAAGATCA 8610

387 GGCACCTCTACCAATGAGAGGCTCTTTTGTGTTGGTTTGTCTCTGATTAAGATCA 328

8611 GGCAGCTGTAATAGTCTAGCGGCTACATAAATCTGCTCAAAAAGCCTATAGAG 8670

327 GGCAGCTGTAATAGTCTAGCGGCTACATAAATCTGCTCAAAAAGCCTATAGAG 268

8671 GTAGGAGGTGAGAGCTTAAGAGAGCCGCTGATATGACACAGAGATAGCC 8730

267 GTAGGAGGTGAGAGCTTAAGAGAGCCGCTGATATGACACAGAGATAGCC 208

8731 TGCACATATATAGCAAGACCTTTGTTCAAAAATGAGAGGAGGATATGTTTAAAGTCT 8790

207 TGCACATATATAGCAAGACCTTTGTTCAAAAATGAGAGGAGGATATGTTTAAAGTCT 148

8791 GGGCTGTGTAAACGACCTAAGGAGGACATGTAGACATTTGACTAAGAAAGATCTCA 8850

147 GGGCTGTGTAAACGACCTAAGGAGGACATGTAGACATTTGACTAAGAAAGATCTCA 88

8851 TCAAAAGCCGGGTGAGGAGGTGAGCTACAGTGTGCTCAAGAGCCCATATAGAGG 8910

87 TCAAAAGCCGGGTGAGGAGGTGAGCTACAGTGTGCTCAAGAGCCCATATAGAGG 28

8911 CAGTTCCCTCTCTCTC 8928

27 CAGTTCCCTCTCTCTC 10

## FEATURES

## SOURCE

Location/Qualifiers

1..495

/organism="Mus musculus"

/mol\_type="genomic DNA"

/strain="C57BL/6J"

/db\_xref="taxon:10090"

/clone="RPCI-24-263015"

/sex="Male"

/cell\_type="Spleen/Brain"

/note="Vector: pPRBAC1; site\_1: BamHI; site\_2: BamHI; RPCI-24 Mouse BAC library produced by Pieter de Jong. The library was cloned in the pPRBAC1 cloning vector at the BamHI sites using MboI partially digested male C57BL/6J DNA."

BASE COUNT 107 a 148 c 153 g 87 t  
ORIGIN

Query Match 5.0%; Score 445.6; DB 28; Length 495;  
Best Local Similarity 97.6%; Pred. No. 5e-38;

Matches 484; Conservative 0; Mismatches 9; Indels 3; Gaps 3;

2740 GGGAGCAGGCGCATGTTTCCA-GAGACCTACAGCAGAGGCGCAAGAGATCCCCAG 2798

1 GGGAGCAGGCGCATGTTTCCAATTTTACCTACAGCAGAGGCGCAAGAGATCCCCAG 60

2799 GTCCAGGCGCAGGAGGTGAGGCGCTTTGTTCCGAGGAGGAGGCGGCGAGCAAGGCT 2858

61 GTCCAGGCGCAGGAGGTGAGGCGCTTTGTTCCGAGGAGGAGGCGGCGAGCAAGGCT 120

2859 TCAAAAGCAGAGGTTTATGACAGCTCAATAAAGTGAGAGTCTGAGGCTCAGTCAAGAAAGGA 2918

121 TCAAAAGCAGAGGTTTATGACAGCTCAATAAAGTGAGAGTCTGAGGCTCAGTCA- TAAGTG 179



Fax: 801 585 7177  
 Email: ddunn@genetics.utah.edu  
 Insert Length: 10000 Std Error: 0.00  
 Plate: 0093 row: E column: 23  
 Seq primer: CACACAGAAACAGCATATGCC  
 Class: plasmid ends  
 High quality sequence stop: 307.

## FEATURES

source

1. 307  
 /organism="Mus musculus"  
 /mol\_type="genomic DNA"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="U09C1M0093E23"  
 /sex="Male"  
 /lab\_host="E. coli strain XL10-Gold, T1-resistant, F-"  
 /clone\_lib="Mouse 10kb plasmid U09C1M library"  
 /note="Vector: PMD42nv, Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of PMD42 (g1147321149b1AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 63 a 67 c 83 g 94 t  
 ORIGIN

Query Match 3.2%; Score 291.8; DB 28; Length 307;  
 Best Local Similarity 99.0%; Pred. No. 1.1e-21;  
 Matches 304; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

QY 1084 GCCAAGCTGGGCTTCTGCAAGC-AGCCATGGCCTTAACCATGGACCGCTCTCT 1142  
 DB 1 GCCAAGCTGGGCTTCTGCAAGC-AGCCATGGCCTTAACCATGGACCGCTCTCT 60  
 QY 1143 AGCCTAAGGTATCTTTAGTTTAAAAATATATATCTCAGCCGGGTGCTGGCAC 1202  
 DB 61 AGCCTAAGGTATCTTTAGTTTAAAAATATATATCTCAGCCGGGTGCTGGCAC 120  
 QY 1203 ACGCCCTTAATCCACACACTTGAGAGCTGAGGTAGAAATTATACACAGCCAGCT 1262  
 DB 121 ACGCCCTTAATCCACACACTTGAGAGCTGAGGTAGAAATTATACACAGCCAGCT 180  
 QY 1263 GGGGTGACAGCTGGCCCTGTTTCTTTCTTTATATGCACTGCTCTTACC 1322  
 DB 181 GGGGTGACAGCTGGCCCTGTTTCTTTCTTTATATGCACTGCTCTTACC 240  
 QY 1323 TGGGTATGTCCTGCAAGGCTGATCCCTTGAGCTGAGTTAAAGACAGTTGTG 1382  
 DB 241 TGGGTATGTCCTGTCAGAGGCTGATCCCTTGAGCTGAGTTAAAGACAGTTGTG 300  
 QY 1383 ATCAGC 1389  
 DB 301 ATCAGC 307

RESULT 5  
 A2821333/c 291 bp DNA linear GSS 20-FEB-2001  
 LOCUS A2821333  
 DEFINITION 2M0094K06F Mouse 10kb plasmid U09C1M library Mus musculus genomic  
 clone U09C2M0094K06 F, genomic survey sequence.

ACCESSION A2821333  
 VERSION A2821333.1 GI:12991241  
 KEYWORDS GSS.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus

REFERENCE  
 AUTHORS 1 (bases 1 to 291)  
 Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,  
 Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly  
 M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.  
 and Wright,D., Weis,R.  
 Mouse whole genome scaffolding with paired end reads from 10kb  
 plasmid inserts

TITLE  
 JOURNAL  
 COMMENT

Unpublished  
 Contact: Robert B. Weiss  
 University of Utah  
 University of Utah  
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SIC, UT  
 84112, USA  
 Tel: 801 585 5606  
 Fax: 801 585 7177  
 Email: ddunn@genetics.utah.edu  
 Insert Length: 10000 Std Error: 0.00  
 Plate: 0094 row: K column: 06  
 Seq primer: CGTTGTAACGACGCGCCACT  
 Class: plasmid ends  
 High quality sequence stop: 291.

## FEATURES

source

1. 291  
 /organism="Mus musculus"  
 /mol\_type="genomic DNA"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="U09C2M0094K06"  
 /sex="Male"  
 /lab\_host="E. coli strain XL10-Gold, T1-resistant, F-"  
 /clone\_lib="Mouse 10kb plasmid U09C1M library"  
 /note="Vector: PMD42nv, Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of PMD42 (g1147321149b1AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 40 a 89 c 77 g 85 t  
 ORIGIN

Query Match 3.1%; Score 280; DB 28; Length 291;  
 Best Local Similarity 99.7%; Pred. No. 2.1e-20;  
 Matches 291; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 2583 ACACACACACACACACACACATTCAGTCTCAGAGCTCTTGGAGGTCACG 2642  
 DB 291 ACACACACACACACACACACATTCAGTCTCAGAGCTCTTGGAGGTCACG 232  
 QY 2643 AAGAGCTGCCCTCAACACAGATCTTATCTTCCCTCTAAAGAGACACAGATTCCAA 2702  
 DB 231 AAGAGGTCGCCCTCAACACAGATCTTATCTTCCCTCTAAAGAGACACAGATTCCAA 172  
 QY 2703 GGTGGCAGAAGATCTACAGGGGCGACAGGAGGAGGAGGAGCAGGCCATGTTCACG 2762







Db	334	GGCTACACAGCCAGCCCTGTCTCTCAGAGATAATATATGCTGGCGGTGGTGGCCGAC	275
Oy	4608	GCCTTTATCCAGCACTTGGGAGGACAGGACGACATTTCTGATTTCAAGCCGACCC	4667
Db	274	GCCTTTATCCAGCACTTGGGAGGACAGGACGACATTTCTGATTTCAAGCCGACCC	215
Oy	4668	TGCTCTATAGAGTATTCACAGACAGCCAGGCTTACAGAGAAACCTCTTTTGAAA	4727
Db	214	TGCTCTACAGAGTATTCACAGAGAGCCAGGACGACGATTTCTGATTTCAAGCCGACCC	155
Oy	4728	ACCGAGAAACACAAACAAACAAACAAACAAACCCAAACCCAA 4774	
Db	154	CCCAAAAAAAAAAAAAAAAAAGAAAGAAAGAAAGAA 108	
RESULT 9			
AK052834/c	4682 bp	mRNA	linear
LOCUS			HTC 05-DEC-2002
DEFINITION			Mus musculus 10 days lactation, adult female mammary gland cDNA, RIKEN full-length enriched library, clone:D730043H12
ACCESSION	AK052834		
VERSION	AK052834.1	GI:26095403	
KEYWORDS	HTC: CAP trapper.		
SOURCE	Mus musculus (house mouse)		
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
REFERENCE			
AUTHORS	1	Carninci, P. and Hayashizaki, Y.	
JOURNAL		High-efficiency full-length cDNA cloning	
MEDLINE		Meth. Enzymol. 303, 19-44 (1999)	
PUBMED	99279253		
2	10349636		
REFERENCE			
AUTHORS	2	Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.	
TITLE		Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes	
JOURNAL		Genome Res. 10 (10), 1617-1630 (2000)	
MEDLINE	20499374		
PUBMED	11042159		
3			
REFERENCE			
AUTHORS	3	Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Kono, H., Akiyama, J., Nishi, K., Kitsuina, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watabiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.	
TITLE		RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer	
JOURNAL		Genome Res. 10 (11), 1757-1771 (2000)	
MEDLINE	20530913		
PUBMED	11076861		
4			
REFERENCE			
AUTHORS		Karai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y., Arakawa, T., Hara, A., Fukunishi, Y., Kono, H., Adachi, T., Fukuda, S., Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamada, I., Saito, T., Okazaki, Y., Gojobori, T., Bono, H., Kasukawa, T., Saito, R., Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T., Fleischmann, W., Gaasterland, T., Gissi, C., Kling, B., Kochiwa, H., Kuehl, P., Lewis, S., Matsuo, Y., Nikaido, I., Pesole, G., Quackenbush, J., Schriml, L. M., Stanbly, F., Suzuki, R., Tomita, M., Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, H., Aono, H., Baldairelli, R., Barsh, G., Blake, J., Boffelli, D., Bojunga, N., Carninci, P., de Bonaldo, M. F., Brownstein, M. J., Bull, C., Fletcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D., Hochmann, M., Hume, D. A., Kamiya, M., Lee, N. H., Lyons, P., Mochizuki, L., Mashima, J., Mazzarelli, J., Mombasers, P., Norioka, P., Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H., Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K. F., Suzuki, H., Toyooka, K., Wang, K. H., Weitz, C., Whitaker, C., Wilming, L.,	

Db	4608	GCCTTTATCCAGCACTTGGGAGGACAGGACGACGATTTCTGATTTCAAGCCGACCC	4667
Oy	4548	AGCATPACAGGAGAGCTTGTCTTAACAAAAATGAAGCCGAGTGGTGGACAC	4607
Db	2087	GCCTACACAGCAGCCCTGTCTCAGAGATAATATATGCTGGCGGTGGTGGCCGAC	2028
Oy	4608	GCCTTTATCCAGCACTTGGGAGGACAGGACGACGATTTCTGATTTCAAGCCGACCC	4667

Db	4608	GCCTTTATCCAGCACTTGGGAGGACAGGACGACGATTTCTGATTTCAAGCCGACCC	4667
Oy	4548	AGCATPACAGGAGAGCTTGTCTTAACAAAAATGAAGCCGAGTGGTGGACAC	4607
Db	2087	GCCTACACAGCAGCCCTGTCTCAGAGATAATATATGCTGGCGGTGGTGGCCGAC	2028
Oy	4608	GCCTTTATCCAGCACTTGGGAGGACAGGACGACGATTTCTGATTTCAAGCCGACCC	4667

Db	4608	GCCTTTATCCAGCACTTGGGAGGACAGGACGACGATTTCTGATTTCAAGCCGACCC	4667
Oy	4548	AGCATPACAGGAGAGCTTGTCTTAACAAAAATGAAGCCGAGTGGTGGACAC	4607
Db	2087	GCCTACACAGCAGCCCTGTCTCAGAGATAATATATGCTGGCGGTGGTGGCCGAC	2028
Oy	4608	GCCTTTATCCAGCACTTGGGAGGACAGGACGACGATTTCTGATTTCAAGCCGACCC	4667

Db	4608	GCCTTTATCCAGCACTTGGGAGGACAGGACGACGATTTCTGATTTCAAGCCGACCC	4667
Oy	4548	AGCATPACAGGAGAGCTTGTCTTAACAAAAATGAAGCCGAGTGGTGGACAC	4607
Db	2087	GCCTACACAGCAGCCCTGTCTCAGAGATAATATATGCTGGCGGTGGTGGCCGAC	2028
Oy	4608	GCCTTTATCCAGCACTTGGGAGGACAGGACGACGATTTCTGATTTCAAGCCGACCC	4667

Db	4608	GCCTTTATCCAGCACTTGGGAGGACAGGACGACGATTTCTGATTTCAAGCCGACCC	4667
Oy	4548	AGCATPACAGGAGAGCTTGTCTTAACAAAAATGAAGCCGAGTGGTGGACAC	4607
Db	2087	GCCTACACAGCAGCCCTGTCTCAGAGATAATATATGCTGGCGGTGGTGGCCGAC	2028
Oy	4608	GCCTTTATCCAGCACTTGGGAGGACAGGACGACGATTTCTGATTTCAAGCCGACCC	4667

Db	4608	GCCTTTATCCAGCACTTGGGAGGACAGGACGACGATTTCTGATTTCAAGCCGACCC	4667
Oy	4548	AGCATPACAGGAGAGCTTGTCTTAACAAAAATGAAGCCGAGTGGTGGACAC	4607
Db	2087	GCCTACACAGCAGCCCTGTCTCAGAGATAATATATGCTGGCGGTGGTGGCCGAC	2028
Oy	4608	GCCTTTATCCAGCACTTGGGAGGACAGGACGACGATTTCTGATTTCAAGCCGACCC	4667

Db	4608	GCCTTTATCCAGCACTTGGGAGGACAGGACGACGATTTCTGATTTCAAGCCGACCC	4667
Oy	4548	AGCATPACAGGAGAGCTTGTCTTAACAAAAATGAAGCCGAGTGGTGGACAC	4607
Db	2087	GCCTACACAGCAGCCCTGTCTCAGAGATAATATATGCTGGCGGTGGTGGCCGAC	2028
Oy	4608	GCCTTTATCCAGCACTTGGGAGGACAGGACGACGATTTCTGATTTCAAGCCGACCC	4667

Db	4608	GCCTTTATCCAGCACTTGGGAGGACAGGACGACGATTTCTGATTTCAAGCCGACCC	4667
Oy	4548	AGCATPACAGGAGAGCTTGTCTTAACAAAAATGAAGCCGAGTGGTGGACAC	4607
Db	2087	GCCTACACAGCAGCCCTGTCTCAGAGATAATATATGCTGGCGGTGGTGGCCGAC	2028
Oy	4608	GCCTTTATCCAGCACTTGGGAGGACAGGACGACGATTTCTGATTTCAAGCCGACCC	4667

Db	4608	GCCTTTATCCAGCACTTGGGAGGACAGGACGACGATTTCTGATTTCAAGCCGACCC	4667
Oy	4548	AGCATPACAGGAGAGCTTGTCTTAACAAAAATGAAGCCGAGTGGTGGACAC	4607
Db	2087	GCCTACACAGCAGCCCTGTCTCAGAGATAATATATGCTGGCGGTGGTGGCCGAC	2028
Oy	4608	GCCTTTATCCAGCACTTGGGAGGACAGGACGACGATTTCTGATTTCAAGCCGACCC	4667

Db	4608	GCCTTTATCCAGCACTTGGGAGGACAGGACGACGATTTCTGATTTCAAGCCGACCC	4667
Oy	4548	AGCATPACAGGAGAGCTTGTCTTAACAAAAATGAAGCCGAGTGGTGGACAC	4607
Db	2087	GCCTACACAGCAGCCCTGTCTCAGAGATAATATATGCTGGCGGTGGTGGCCGAC	2028
Oy	4608	GCCTTTATCCAGCACTTGGGAGGACAGGACGACGATTTCTGATTTCAAGCCGACCC	4667

Db	4608	GCCTTTATCCAGCACTTGGGAGGACAGGACGACGATTTCTGATTTCAAGCCGACCC	4667
Oy	4548	AGCATPACAGGAGAGCTTGTCTTAACAAAAATGAAGCCGAGTGGTGGACAC	4607
Db	2087	GCCTACACAGCAGCCCTGTCTCAGAGATAATATATGCTGGCGGTGGTGGCCGAC	2028
Oy	4608	GCCTTTATCCAGCACTTGGGAGGACAGGACGACGATTTCTGATTTCAAGCCGACCC	4667

Db	4608	GCCTTTATCCAGCACTTGGGAGGACAGGACGACGATTTCTGATTTCAAGCCGACCC	4667
Oy	4548	AGCATPACAGGAGAGCTTGTCTTAACAAAAATGAAGCCGAGTGGTGGACAC	4607
Db	2087	GCCTACACAGCAGCCCTGTCTCAGAGATAATATATGCTGGCGGTGGTGGCCGAC	2028
Oy	4608	GCCTTTATCCAGCACTTGGGAGGACAGGACGACGATTTCTGATTTCAAGCCGACCC	4667

Db	4608	GCCTTTATCCAGCACTTGGGAGGACAGGACGACGATTTCTGATTTCAAGCCGACCC	4667
Oy	4548	AGCATPACAGGAGAGCTTGTCTTAACAAAAATGAAGCCGAGTGGTGGACAC	4607
Db	2087	GCCTACACAGCAGCCCTGTCTCAGAGATAATATATGCTGGCGGTGGTGGCCGAC	2028
Oy	4608	GCCTTTATCCAGCACTTGGGAGGACAGGACGACGATTTCTGATTTCAAGCCGACCC	4667

Db	4608	GCCTTTATCCAGCACTTGGGAGGACAGGACGACGATTTCTGATTTCAAGCCGACCC	4667
Oy	4548	AGCATPACAGGAGAGCTTGTCTTAACAAAAATGAAGCCGAGTGGTGGACAC	4607
Db	2087	GCCTACACAGCAGCCCTGTCTCAGAGATAATATATGCTGGCGGTGGTGGCCGAC	2028
Oy	4608	GCCTTTATCCAGCACTTGGGAGGACAGGACGACGATTTCTGATTTCAAGCCGACCC	4667

Db	4608	GCCTTTATCCAGCACTTGGGAGGACAGGACGACGATTTCTGATTTCAAGCCGACCC	4667
Oy	4548	AGCATPACAGGAGAGCTTGTCTTAACAAAAATGAAGCCGAGTGGTGGACAC	4607
Db	2087	GCCTACACAGCAGCCCTGTCTCAGAGATAATATATGCTGGCGGTGGTGGCCGAC	2028
Oy	4608	GCCTTTATCCAGCACTTGGGAGGACAGGACGACGATTTCTGATTTCAAGCCGACCC	4667

Db	4608	GCCTTTATCCAGCACTTGGGAGGACAGGACGACGATTTCTGATTTCAAGCCGACCC	4667
Oy	4548	AGCATPACAGGAGAGCTTGTCTTAACAAAAATGAAGCCGAGTGGTGGACAC	4607
Db	2087	GCCTACACAGCAGCCCTGTCTCAGAGATAATATATGCTGGCGGTGGTGGCCGAC	2028
Oy	4608	GCCTTTATCCAGCACTTGGGAGGACAGGACGACGATTTCTGATTTCAAGCCGACCC	4667

Db	4608	GCCTTTATCCAGCACTTGGGAGGACAGGACGACGATTTCTGATTTCAAGCCGACCC	4667
Oy	4548	AGCATPACAGGAGAGCTTGTCTTAACAAAAATGAAGCCGAGTGGTGGACAC	4607
Db	2087	GCCTACACAGCAGCCCTGTCTCAGAGATAATATATGCTGGCGGTGGTGGCCGAC	2028
Oy	4608	GCCTTTATCCAGCACTTGGGAGGACAGGACGACGATTTCTGATTTCAAGCCGACCC	4667

Db	4608	GCCTTTATCCAGCACTTGGGAGGACAGGACGACGATTTCTGATTTCAAGCCGACCC	4667
Oy	4548	AGCATPACAGGAGAGCTTGTCTTAACAAAAATGAAGCCGAGTGGTGGACAC	4607
Db	2087	GCCTACACAGCAGCCCTGTCTCAGAGATAATATATGCTGGCGGTGGTGGCCGAC	2028
Oy	4608	GCCTTTATCCAGCACTTGGGAGGACAGGACGACGATTTCTGATTTCAAGCCGACCC	4667

Db	4608	GCCTTTATCCAGCACTTGGGAGGACAGGACGACGATTTCTGATTTCAAGCCGACCC	4667
Oy	4548	AGCATPACAGGAGAGCTTGTCTTAACAAAAATGAAGCCGAGTGGTGGACAC	4607
Db	2087	GCCTACACAGCAGCCCTGTCTCAGAGATAATATATGCTGGCGGTGGTGGCCGAC	2028
Oy	4608	GCCTTTATCCAGCACTTGGGAGGACAGGACGACGATTTCTGATTTCAAGCCGACCC	4667

Db	4608	GCCTTTATCCAGCACTTGGGAGGACAGGACGACGATTTCTGATTTCAAGCCGACCC	4667
Oy	4548	AGCATPACAGGAGAGCTTGTCTTAACAAAAATGAAGCCGAGTGGTGGACAC	4607
Db	2087	GCCTACACAGCAGCCCTGTCTCAGAGATAATATATGCTGGCGGTGGTGGCCGAC	2028
Oy	4608	GCCTTTATCCAGCACTTGGGAGGACAGGACGACGATTTCTGATTTCAAGCCGACCC	4667

Db	4608	GCCTTTATCCAGCACTTGGGAGGACAGGACGACGATTTCTGATTTCAAGCCGACCC	4667
Oy	4548	AGCATPACAGGAGAGCTTGTCTTAACAAAAATGAAGCCGAGTGGTGGACAC	4607
Db	2087	GCCTACACAGCAGCCCTGTCTCAGAGATAATATATGCTGGCGGTGGTGGCCGAC	2028
Oy	4608	GCCTTTATCCAGCACTTGGGAGGACAGGACGACGATTTCTGATTTCAAGCCGACCC	4667

Db	4608	GCCTTTATCCAGCACTTGGGAGGACAGGACGACGATTTCTGATTTCAAGCCGACCC	4667
Oy	4548	AGCATPACAGGAGAGCTTGTCTTAACAAAAATGAAGCCGAGTGGTGGACAC	4607
Db	2087	GCCTACACAGCAGCCCTGTCTCAGAGATAATATATGCTGGCGGTGGTGGCCGAC	2028
Oy	4608	GCCTTTATCCAGCACTTGGGAGGACAGGACGACGATTTCTGATTTCAAGCCGACCC	4667

Db	4608	GCCTTTATCCAGCACTTGGGAGGACAGGACGACGATTTCTGATTTCAAGCCGACCC	4667
Oy	4548	AGCATPACAGGAGAGCTTGTCTTAACAAAAATGAAGCCGAGTGGTGGACAC	4607
Db	2087	GCCTACACAGCAGCCCTGTCTCAGAGATAATATATGCTGGCGGTGGTGGCCGAC	2028
Oy	4608	GCCTTTATCCAGCACTTGGGAGGACAGGACGACGATTTCTGATTTCAAGCCGACCC	4667

Db	4608	GCCTTTATCCAGCACTTGGGAGGACAGGACGACGATTTCTGATTTCAAGCCGACCC	4667
Oy	4548	AGCATPACAGGAGAGCTTGTCTTAACAAAAATGAAGCCGAGTGGTGGACAC	4607
Db	2087	GCCTACACAGCAGCCCTGTCTCAGAGATAATATATGCTGGCGGTGGTGGCCGAC	2028
Oy	4608	GCCTTTATCCAGCACTTGGGAGGACAGGACGACGATTTCTGATTTCAAGCCGACCC	4667

Db	4608	GCCTTTATCCAGCACTTGGGAGGACAGGACGACGATTTCTGATTTCAAGCCGACCC	4667
Oy	4548	AGCATPACAGGAGAGCTTGTCTTAACAAAAATGAAGCCGAGTGGTGGACAC	4607
Db	2087	GCCTACACAGCAGCCCTGTCTCAGAGATAATATATGCTGGCGGTGGTGGCCGAC	2028
Oy	4608	GCCTTTATCCAGCACTTGGGAGGACAGGACGACGATTTCTGATTTCAAGCCGACCC	4667

Db	4608	GCCTTTATCCAGCACTTGGGAGGACAGGACGACGATTTCTGATTTCAAGCCGACCC	4667
Oy	4548	AGCATPACAGGAGAGCTTGTCTTAACAAAAATGAAGCCGAGTGGTGGACAC	4607
Db	2087	GCCTACACAGCAGCCCTGTCTCAGAGATAATATATGCTGGCGGTGGTGGCCGAC	2028
Oy	4608	GCCTTTATCCAGCACTTGGGAGGACAGGACGACGATTTCTGATTTCAAGCCGACCC	4667

Db	4608	GCCTTTATCCAGCACTTGGGAGGACAGGACGACGATTTCTGATTTCAAGCCGACCC	4667
Oy	4548	AGCATPACAGGAGAGCTTGTCTTAACAAAAATGAAGCCGAGTGGTGGACAC	4607
Db	2087	GCCTACACAGCAGCCCTGTCTCAGAGATAATATATGCTGGCGGTGGTGGCCGAC	2028
Oy	4608	GCCTTTATCCAGCACTTGGGAGGACAGGACGACGATTTCTGATTTCAAGCCGACCC	4667

Db	4608	GCCTTTATCCAGCACTTGGGAGGACAGGACGACGATTTCTGATTTCAAGCCGACCC	4667
Oy	4548	AGCATPACAGGAGAGCTTGTCTTAACAAAAATGAAGCCGAGTGGTGGACAC	4607
Db	2087	GCCTACACAGCAGCCCTGTCTCAGAGATAATATATGCTGGCGGTGGTGGCCGAC	2028
Oy	4608	GCCTTTATCCAGCACTTGGGAGGACAGGACGACGATTTCTGATTTCAAGCCGACCC	4667

Db	4608	GCCTTTATCCAGCACTTGGGAGGACAGGACGACGATTTCTGATTTCAAGCCGACCC	4667
Oy	4548	AGCATPACAGGAGAGCTTGTCTTAACAAAAATGAAGCCGAGTGGTGGACAC	4607
Db	2087	GCCTACACAGCAGCCCTGTCTCAGAGATAATATATGCTGGCGGTGGTGGCCGAC	2028
Oy	4608	GCCTTTATCCAGCACTTGGGAGGACAGGACGACGATTTCTGATTTCAAGCCGACCC	4667

Db	4608	GCCTTTATCCAGCACTTGGGAGGACAGGACGACGATTTCTGATTTCAAGCCGACCC	4667
Oy	4548	AGCATPACAGGAGAGCTTGTCTTAACAAAAATGAAGCCGAGTGGTGGACAC	4607
Db	2087	GCCTACACAGCAGCCCTGTCTCAGAGATAATATATGCTGGCGGTGGTGGCCGAC	2028
Oy	4608	GCCTTTATCCAGCACTTGGGAGGACAGGACGACGATTTCTGATTTCAAGCCGACCC	4667

Db	4608	GCCTTTATCCAGCACTTGGGAGGACAGGACGACGATTTCTGATTTCAAGCCGACCC	4667
Oy	4548	AGCATPACAGGAGAGCTTGTCTTAACAAAAATGAAGCCGAGTGGTGGACAC	4607
Db	2087	GCCTACACAGCAGCCCTGTCTCAGAGATAATATATGCTGGCGGTGGTGGCCGAC	2028
Oy	4608	GCCTTTATCCAGCACTTGGGAGGACAGGACGACGATTTCTGATTTCAAGCCGACCC	4667

Db	4608	GCCTTTATCCAGCACTTGGGAGGACAGGACGACGATTTCTGATTTCAAGCCGACCC	4667
Oy	4548	AGCATPACAGGAGAGCTTGTCTTAACAAAAATGAAGCCGAGTGGTGGACAC	4607
Db	2087	GCCTACACAGCAGCCCTGTCTCAGAGATAATATATGCTGGCGGTGGTGGCCGAC	2028
Oy	4608	GCCTTTATCCAGCACTTGGGAGGACAGGACGACGATTTCTGATTTCAAGCCGACCC	4667

Db	4608	GCCTTTATCCAGCACTTGGGAGGACAGGACGACGATTTCTGATTTCAAGCCGACCC	4667
Oy	4548	AGCATPACAGGAGAGCTTGTCTTAACAAAAATGAAGCCGAGTGGTGGACAC	4607
Db	2087	GCCTACACAGCAGCCCTGTCTCAGAGATAATATATGCTGGCGGTGGTGGCCGAC	2028
Oy	4608	GCCTTTATCCAGCACTTGGGAGGACAGGACGACGATTTCTGATTTCAAGCCGACCC	4667

Db	4608	GCCTTTATCCAGCACTTGGGAGGACAGGACGACGATTTCTGATTTCAAGCCGACCC	466
----	------	--	-----

[illegible]

OY		4489	TAACTGTGACCTTGGGGAGCGCTGAAGAG-G-CATTTCCTATATGTTTGAGCCAGCCTG	454.7
Db		323	TATTCCCAAGCAGCTGGGAAGCTGAGACAAGACTTAACCTAAGAAGTTCAAGCTTAGCCTG	264
OY		4548	AGCTATAGAGGAGCACTTGTGCTTTAAGAAAAATAATGAAGCCCAAGCAGTGGTGGCACAC	4607
Db		263	GGCTACACAGCAGGCCCTGTCTCAGAGAAATAATATGCTGTGGCGGCTGTGGCGCAC	204
OY		4608	GCCCTTAATCCCAGCACTTGGGAGGCAGAACAGCAGATTTCTGTAGTTCAAGGCCAGCC	4667
Db		203	GCCCTTAATCCACACACTTGGGAGGCAGAACAGCAGTGTGATCTGTAGTTCCAGGCCAGCC	144
OY		4668	TGGCTCTATAGAGTAGTTCACAGGACAGCAGCGGTACACAGAAAACCCTTTTGGAAAA	4727
Db		143	TGGCTCTACAGAGTGTGCTTCACAGACAGCCAGCAAGACTACACAGAAACCCTTTTGGAAAA	84
OY		4728	ACCAGAAAAACAAACAAACAAACAAACAAA 4761	
Db		83	CCCCAAAAAAAAAAAAAAAAAGAAAGAAAAAAA 50	
RESULT 11				
B1465446/c				
DEFINITION		B1465446	399 bp mRNA linear EST 22-AUG-2001	
ACCESSION		U167C03.Y1	Kaestner ngn3 wt Mus musculus cDNA 5', mRNA sequence.	
VERSION		B1465446.1	GI:15277953	
KEYWORDS		EST.		
SOURCE		Mus musculus (house mouse)		
ORGANISM		Mus musculus		
REFERENCE		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS		Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
JOURNAL		1 (bases 1 to 399)		
COMMENT		Melton,D., Brown,J., Kenty,G., Permult,A., Lee,C., Kaestner,K., Lemishka,I., Searce,M., Brestelli,J., Gradwohl,G., Clifton,S., Hillier,L., Marra,M., Pape,D., Wylie,T., Martin,J., Blistain,A., Schmitt,A., Theising,B., Ritter,E., Ronko,I., Bennett,J., Cardenas ,M., Gibbons,M., McMan,R., Cole,R., Tsagariswhill,R., Williams,T., Jackson,Y. and Bowers,Y.		
TITLE		Endocrine Pancreas Consortium		
UNPUBLISHED		Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue		
ENDOCRINE PANCREAS CONSORTIUM		Harvard University, Howard Hughes Medical Institute		
DEPT OF MOLECULAR AND CELLULAR BIOLOGY, 7 DIVINITY AVE, CAMBRIDGE,		MA 02138		
TEL: 617-495-1812				
FAX: 617-495-8557				
Email: dmelton@lhop.harvard.edu				
Pancreas was obtained from Gerard Gradwohl (PMAS 97 P1607-1611,				
2000) Library was constructed by Catherine Lee DNA sequencing by:				
Washington University Genome Sequencing Center for information on				
obtaining a clone please contact: Dr. Marie Searce				
(mscarce@mail.med.upenn.edu)				
High quality sequence stop: 268.				
Location/Qualifiers				
1..399				
/organism="Mus musculus"				
/mol_type="mRNA"				
/strain="129/Sv x CD1"				
/db_xref="taxon:10090"				
/dev_stage="p.c. 14.5"				
/lab_host="E. coli-DH12S (GIBCO)"				
/clone_id="Kaestner ngn3 wt"				
/note="Organ: pancreas; Vector: pSPOR11 (GIBCO); Site_1:				
Not I; Site_2: Sal I; The library was prepared by				
Catherine S. Lee and has not been published. The pancreas				
was obtained from Gerard Gradwohl (PMAS 97 P1607-1611,				
2000). The cDNA's were prepared with an oligo containing a				
NotI site, and SalI linkers were added to the ends. The				
inserts were cut with NotI before being cloned into the				
Not-SalI sites in the vectors. This is one of two				

libraries, ngn3 wt and ngn3 -/- . The wt library is in  
psPORT1, T7 promoter is 5'."

BASE COUNT 86 a 99 c 79 g 135 t  
ORIGIN

Query Match 1.9%; Score 173.4; DB 12; Length 399;

Best Local Similarity 79.3%; Pred. No. 3.3e-09;

Matches 218; Conservative 0; Mismatches 56; Indels 1; Gaps 1;

4489 TAATCTGAGCACTTGGAGGCGAGGAG-GATTGCTATATGTTGAGCCACCTCG 4547  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
302 TATTCACAGAGCTGGAGGCTGAGAGCAAGTATCAATAGAAAGTTCAGAGGCTAGCCTG 243  
4548 AGCTATAGAGCAGACTTGTCTTAAAGAAAAATGAAGCCAGAGTGTGCGACAC 4607  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
242 GGCTACAGAGCGAGCCCTGTCTCAGAGATTAATTAATGCTGGCGGTGTGCGCAC 183  
4608 GCCTTAAATCCAGCACTTGGAGGCGAGAGCGAGATTTCTGAGTTCAAGGCGAGCC 4667  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
182 GCCTTAAATCCAGCACTTGGAGGCGAGAGCGAGGATCTCTGAGTTCAAGGCGAGCC 123  
4668 TGGCTATAGAGTGAGTTCACAGAGCGGCTACACAGAAACCTGTTTGAAAA 4727  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
122 TGGTCTACAGAGTGAGTTCACAGAGCGGAGCTACAGAGAAACCTGTTTGAAAA 63  
4728 ACCAGAAAAACAAAAACAAAAACAAAAACAAAA 4762  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
62 CCAAAAAAAGAAAAAAGAAAAAAGAAAAA 28

## RESULT 12

BM875096 383 bp mRNA linear EST 07-MAR-2002

LOCUS BM875096.1 GI:19242755

DEFINITION 1f37f06.y1 Kaestner wt amplified Mus musculus CDNA 5', mRNA

Sequence.

ACCESSION BM875096

VERSION BM875096.1 GI:19242755

KEYWORDS EST.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus;

1 (bases 1 to 383)

REFERENCE Melton, D., Brown, J., Kenty, G., Permutt, A., Lee, C., Kaestner, K.,

Hillier, L., Marra, M., Pape, D., Wylie, T., Martin, J., Bistrain, A.,

Schmitt, A., Theising, B., Rittler, E., Ronko, I., Bennett, J., Cadenas,

M., Gibbons, M., McCann, R., Cole, R., Tsagaris, R., Williams, T.,

Jackson, Y., and Bowers, Y.

Endocrine Pancreas Consortium

Unpublished

CONTACT: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue

Harvard University, Howard Hughes Medical Institute

Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,

MA 02138

Tel: 617-495-1812

Fax: 617-495-8557

Email: dmelton@biohp.harvard.edu

Pancreas was obtained from Gerard Gradwohl (PNAS 97 P1607-1611,

2000) Library was constructed by Catherine Lee DNA sequencing by:

Washington University Genome Sequencing Center for information on

obtaining a clone please contact: Dr. Marie Searce

(mscarce@mail.med.upenn.edu)

Seq primer: -40RP from Gibco

High quality sequence stop: 367.

Location/Qualifiers

1. 383

/organism="Mus musculus"

/mol\_type="mRNA"

/strain="129/Sv x CD1"

/db\_xref="taxon:10090"

/sex="not known"

FEATURES

SOURCE

ORIGIN

/tissue\_type="pancreas"  
/dev\_stage="E14.5"  
/lab\_host="DH12S"  
/clone\_lib="Kaestner wt amplified"

/note="Vector: psPORT1 (GIBCO). Site\_1: Not 1; Site\_2: Sal  
I; The SMARTech PCR cDNA synthesis kit (CLONTECH) was used  
to obtain full-length cDNA from E14.5 wildtype mRNA. A  
modified 3' oligo(dT) containing a NotI site was used to  
prime the first-strand synthesis reaction to allow  
directional cloning. This full-length cDNA was used for  
PCR amplification to generate ds cDNA. After size  
selection (>1kb), SalI adapters were ligated onto both  
ends of the control cDNA, digested with NotI and cloned  
into psPORT1 vector (LTI). The library was electroporated  
into DH12S cells (LTI) and semi-solid amplified. E14.5  
pancreatic primordium was a gift from Dr. Gradwohl."

BASE COUNT 77 a 96 c 81 g 129 t

Query Match 1.9%; Score 172.4; DB 12; Length 383;

Best Local Similarity 79.2%; Pred. No. 4.3e-09;

Matches 217; Conservative 0; Mismatches 56; Indels 1; Gaps 1;

4489 TAATCTGAGCACTTGGAGGCTGAAGCAGAG-GATTGCTATATGTTGAGCCACCTCG 4547  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
289 TATTCACAGAGCTGGAGGCTGAGAGCAAGATTAATAGAAAGTTCAGAGGCTAGCCTG 240  
4548 AGCTATAGAGCAGACTTGTCTTAAAGAAAAATGAAGCCAGAGTGTGCGACAC 4607  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
239 GGCTACAGAGCGAGCCCTGTCTCAGAGATTAATTAATGCTGGCGGTGTGCGCAC 180  
4608 GCCTTAAATCCAGCACTTGGAGGCGAGAGCGAGAGCTTCTGAGTTCAAGGCGAGCC 4667  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
179 GCCTTAAATCCAGCACTTGGAGGCGAGAGCGAGGATCTCTGAGTTCAAGGCGAGCC 120  
4668 TGGCTATAGAGTGAGTTCACAGAGCGGCTACACAGAGAAACCTGTTTGAAAA 4727  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
119 TGGTCTACAGAGTGAGTTCACAGAGCGGAGCTACACAGAGAAACCTGTTTGAAAA 60  
4728 ACCAGAAAAACAAAAACAAAAACAAAAACAAAA 4761  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
59 CCAAAAAAAGAAAAAAGAAAAAAGAAAAA 26

## RESULT 13

AV307409 747 bp mRNA linear EST 24-OCT-2001

LOCUS AV307409

DEFINITION AV307409 RIKEN full-length enriched, 8 days embryo Mus musculus

CDNA clone 5730552H01 3', mRNA sequence.

ACCESSION AV307409

VERSION AV307409.2 GI:16393934

KEYWORDS EST.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus;

1 (bases 1 to 747)

REFERENCE Arakawa, T., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Harz, A.,

Hiramoto, K., Horii, F., Ishii, Y., Ito, M., Kawai, J., Konno, H., Kouda,

M., Koya, S., Matsuyama, T., Miyazaki, A., Nomura, K., Ono, M.,

Okazaki, Y., Okido, T., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki,

D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Suzuki, H.,

Tagami, M., Tagawa, A., Takahashi, F., Takeda, Y., Tanaka, T., Toya, T.,

Muramatsu, M. and Hayashizaki, Y.

RIKEN Mouse ESTs (Arakawa, T., et al. 2001)

Unpublished

CONTACT: Yoshihide Hayashizaki

On Nov 10, 1999 this sequence version replaced gi:6339923.

Laboratory for Genome Exploration Research Group, RIKEN Genomic

Sciences Center (GSC), Yokohama Institute

The Institute of Physical and Chemical Research (RIKEN)

1-7-22 Suenho-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

Tel: 81-45-503-9222







BASE COUNT 71 a 88 c 78 g 109 t  
ORIGIN

Query Match 1.9%; Score 166.8; DB 13; Length 346;  
Best Local Similarity 79.8%; Pred. No. 1.8e-08;  
Matches 209; Conservative 0; Mismatches 52; Indels 1; Gaps 1;

QY 4489 TAATCTGAGCAGCTTGGAGGCTGAAGCAGAGAG-GATTGCTATATGTTTGAAGCCAGCCTG 4547  
|| || |||| |||| |||| || || || || |||| ||||  
Db 287 TATTCCGACAGAGCTGGGAGAGCTGAGACAGAGTTACCATTAAGAGTTCAAGGCTAGCCTG 228  
QY 4548 AGCTATAGAGGAGAGCTTGTCTTTAGAAAAAATGAAGCCAGCAGTGGTGACAC 4607  
|||| |||| |||| |||| |||| |||| |||| |||| ||||  
Db 227 GGCTACACAGCCAGCCCTGTCTCAGAGAAATAATGCTGGCGGTGCTGCGCAC 168  
QY 4608 GCCCTTAATCCAGCAGCTTGGAGGAGAGAGAGAGATTTCTGAGTCAAGGCCAGCC 4667  
|||| |||| |||| |||| |||| |||| |||| |||| ||||  
Db 167 GCCCTTAATCCAGCAGCTTGGAGGAGAGAGAGAGTGTGATCTGAGTTCAGAGGCCAGCC 108  
QY 4668 TGGCTCTATAGAGTGAAGTCCAGAGACAGCCAGGCTACACAGAGAAACCTGTTTGA AAA 4727  
|||| |||| |||| |||| |||| |||| |||| |||| ||||  
Db 107 TGGCTCTACAGAGTGAAGTTCAGAGAGAGGAGACTACACAGAGAAACCTGTTTGA AAA 48  
QY 4728 ACCAGAAAAACAAAAACAACA 4749  
||| |||| |||| |||| ||  
Db 47 ACCCAAAAAAAAAAAAAAAAAA 26

Search completed: September 14, 2003, 09:09:41  
Job time : 16412 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 13, 2003, 18:15:33 ; Search time 438 Seconds  
(without alignments)  
9064.478 Million cell updates/sec

Title: US-09-877-935-1  
Perfect score: 8995  
Sequence: 1 gatctgctgcaccaagaca.....ctctagctcgtccacatg 8995

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 10%  
Listing first 45 summaries

Database : Issued\_Patents\_NA:\*

1: /cgn2-6/ptodata/2/ina/5A.COMB.seq:\*  
2: /cgn2-6/ptodata/2/ina/5B.COMB.seq:\*  
3: /cgn2-6/ptodata/2/ina/6A.COMB.seq:\*  
4: /cgn2-6/ptodata/2/ina/6B.COMB.seq:\*  
5: /cgn2-6/ptodata/2/ina/PTUUS.COMB.seq:\*  
6: /cgn2-6/ptodata/2/ina/backfile1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	148.8	1.7	7208	3	US-09-166-186-107
2	148.8	1.7	7208	3	US-09-313-932-107
3	142.6	1.6	14707	4	US-09-312-762A-3
4	140.4	1.6	3481	3	US-08-965-729A-1
5	138.2	1.5	17056	3	US-09-245-041-3
6	138	1.5	5973	3	US-09-245-041-4
7	136.8	1.5	2509	4	US-09-319-284-1
8	133.4	1.5	90050	3	US-09-245-041-5
9	133.2	1.5	4698	1	US-07-807-043B-5
10	133.2	1.5	4698	1	US-08-299-849B-5
11	133.2	1.5	4698	2	US-08-142-368A-5
12	133.2	1.5	4698	3	US-08-967-727-5
13	133.2	1.5	4698	3	US-08-037-230D-5
14	133.2	1.5	4698	4	US-09-583-850-5
15	133.2	1.5	4698	4	US-09-579-197-5
16	133.2	1.5	4698	4	US-09-404-026-5
17	134	1.5	6727	3	US-08-629-643A-5
18	134	1.5	6727	3	US-09-280-799-1
19	134	1.5	6727	3	US-09-155-884-5
20	133.4	1.5	51259	3	US-08-781-891-209
21	133.4	1.5	51259	4	US-09-618-166-209
22	133.8	1.5	2574	4	US-09-668-313A-10
23	133.8	1.5	17056	3	US-09-245-041-3
24	133.2	1.5	48974	3	US-08-920-422-17
25	131.8	1.5	4550	3	US-09-338-907-182
26	131.8	1.5	4550	3	US-09-218-207-182
27	131.4	1.5	3892	2	US-08-555-723B-3

28	131.4	1.5	3892	3	US-09-123-465-3	Sequence 3, Appl1
29	130.8	1.5	90050	3	US-09-245-041-5	Sequence 5, Appl1
30	130.6	1.5	37950	4	US-09-338-907-183	Sequence 183, App
31	130.6	1.5	37950	4	US-09-218-207-183	Sequence 183, App
32	129.6	1.4	35828	4	US-09-449-218D-17	Sequence 17, Appl
33	129.6	1.4	35828	4	US-09-668-529A-17	Sequence 17, Appl
34	129.6	1.4	35828	4	US-09-312-762A-17	Sequence 17, Appl
35	129.2	1.4	14707	4	US-09-312-762A-3	Sequence 3, Appl1
36	129	1.4	7874	4	US-09-780-175-96	Sequence 96, Appl
37	127.8	1.4	48974	3	US-08-920-422-17	Sequence 17, Appl
38	126.4	1.4	3217	3	US-09-232-200-64	Sequence 64, Appl
39	126.4	1.4	3217	4	US-09-232-197-64	Sequence 64, Appl
40	126.4	1.4	3217	4	US-09-232-201-64	Sequence 64, Appl
41	125.6	1.4	5687	2	US-08-380-403A-3	Sequence 3, Appl1
42	125.6	1.4	5687	2	US-08-895-628-3	Sequence 3, Appl1
43	125.6	1.4	5687	4	US-08-895-810D-3	Sequence 3, Appl1
44	123.6	1.4	1355	1	US-07-757-390-4	Sequence 4, Appl1
45	123.6	1.4	1355	1	US-07-757-390-16	Sequence 16, Appl

ALIGNMENTS

RESULT 1  
US-09-166-186-107  
Sequence 107, Application US/0916186A  
Patent No. 6080580  
GENERAL INFORMATION:  
APPLICANT: Baker, Brenda  
APPLICANT: Bennett, C. Frank  
APPLICANT: Butler, Madeline M.  
APPLICANT: Shanahan, William R.  
TITLE OF INVENTION: ANTISENSE OLIGONUCLEOTIDE MODULATION OF TNF- $\alpha$  EXPRESSION  
FILE REFERENCE: ISPH-0322  
CURRENT APPLICATION NUMBER: US/09/166,186A  
CURRENT FILING DATE: 1998-10-05  
NUMBER OF SEQ ID NOS: 250  
SEQ ID NO 107  
LENGTH: 7208  
TYPE: DNA  
ORGANISM: Mus musculus  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (4527..4712,5225..5279,5457..5504,5799..6217)  
FEATURE:  
NAME/KEY: exon  
LOCATION: (4371)..(4712)  
FEATURE:  
NAME/KEY: Intron  
LOCATION: (4713)..(5224)  
FEATURE:  
NAME/KEY: exon  
LOCATION: (5225)..(5279)  
FEATURE:  
NAME/KEY: Intron  
LOCATION: (5280)..(5456)  
FEATURE:  
NAME/KEY: exon  
LOCATION: (5457)..(5504)  
FEATURE:  
NAME/KEY: Intron  
LOCATION: (5505)..(5798)  
FEATURE:  
NAME/KEY: exon  
LOCATION: (5799)..(6972)  
PUBLICATION INFORMATION:  
AUTHORS: Semon, D.  
AUTHORS: Kawashima, E.  
AUTHORS: Jongeneel, C.V.  
AUTHORS: Shakhov, A.N.  
AUTHORS: Nedospasov, S.A.  
TITLE: Nucleotide sequence of the murine TNF locus, including the  
TITLE: TNF-alpha (tumor necrosis factor) and TNF-beta (lymphotoxin) genes



```
TELEPHONE: 972-3-5625553
TELEFAX: 972-3-5625554
TELEX:
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 14707
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-09-312-762A-3

Query Match
Best Local Similarity 1.6%; Score 142.6; DB 4; Length 14707;
Matches 178; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

QY 4576 AAAAAATGAAAGCCAGAGTGGGACACAGCCTTAATCCCGACCTGGGAGCAG 4635
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 12047 AAGAAATTAAAGCGGGGGGGGATGATGACAGCCTTTATATCCAGCTGGAGCAG 12106
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 4636 AAGCAGCAGATTTCGATTCAGCCAGCAGCTGTATAGAGTTCAGAGCAGC 4695
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 12107 AGCGGGGGGATTTCTGAGTTGAGCCAGCAGCTGTATACAAAGAGTTCAGAGCAGC 12166
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 4696 CAGGCTACACAGAGAAACCTGTTTGAAGAAACAGAAACAAACAAACAA 4755
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 12167 CAGGCTACACAGAGAAACCTGTTTGAAGAAACAAACAAACAAACAA 12226
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 4756 AACAAGCCAAACCAACCAACCTCATCTCTAGCTGTGCT 4812
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 12227 GTTGAAGCAAGTGAAGGCTCCCTTCTGCTGAGCTGAGTTGGGTTT 12283
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 4
US-08-965-729A-1/c
Sequence 1, Application US/08965729A
Patent No. 6200751
GENERAL INFORMATION:
APPLICANT: Jian-Ming Gu and Charles T. Esmon
TITLE OF INVENTION: ENDOTHELIAL SPECIFIC EXPRESSION
TITLE OF INVENTION: REGULATED BY EPOR CONTROL ELEMENTS
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESS: Patrea L. Pabst
STREET: 2800 One Atlantic Center, 1201 West Peachtree Street
CITY: Atlanta
STATE: GA
COUNTRY: USA
ZIP: 30309-4530
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/965,729A
FILING DATE: 07-NOV-1997
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/030,718
FILING DATE: 08-NOV-1997
ATTORNEY/AGENT INFORMATION:
NAME: Pabst, Patrea L.
REGISTRATION NUMBER: 31,284
REFERENCE/DOCKET NUMBER: OMR# 164 PCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: 404-873-8794
TELEFAX: 404-873-8795
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3481 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
```

```
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
FEATURE:
NAME/KEY: misc.feature
OTHER INFORMATION: /note="Nucleotides 2270 through 2840 are a
OTHER INFORMATION: large endothelial specific element"; murine
FEATURE:
NAME/KEY: misc.feature
OTHER INFORMATION: /note="Nucleotides 2990 through 3061 are a
OTHER INFORMATION: serum response element"; murine
FEATURE:
NAME/KEY: misc.feature
OTHER INFORMATION: /note="Nucleotides 3007 through 3014 are a
OTHER INFORMATION: thrombin responsive element"; murine
FEATURE:
NAME/KEY: misc.feature
OTHER INFORMATION: /note="Nucleotides 3130 through 3350 are an
OTHER INFORMATION: endothelial specific element"; murine
US-08-965-729A-1

Query Match
Best Local Similarity 1.6%; Score 140.4; DB 3; Length 3481;
Matches 202; Conservative 0; Mismatches 66; Indels 2; Gaps 2;

QY 4564 TTTGCTTTAAGAAAAATGAAGCCAGCAGTGTGGCAGACGCTTAATCCAGCA 4623
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2563 TTCACTTTAAAAAAGCCGCGGTGTGTGACAGCCTTAATCCAGCA 2504
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 4624 CTGGAGAGCAGACAGCAGATTCTGAGTTCAAGCCAGCTGTATAGAGTAG 4683
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2503 CTGGAGAGCAGACAGCAGATTCTGAGTTCAAGCCAGCTGTATAGAGTAG 2444
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 4684 TTCCAGGAGCAGCAGGCTACACAGAAACCTGTTGAAGAAACCAAA 4742
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2443 -TCCAGAGACAGCAGGCTACACAGAAACCTGTTCAAAACAAACAAACAA 2385
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 4743 CAAACAAACAAACAAACCAACCAACCAACCTCTCATCTCTCTCTA 4802
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2384 AAAAAACAAACAAACCAACCAACCAACCAACCAACCAACCAACCA 2325
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 4803 GCGTGTCTGTCTAGGTGAGTTGG 4832
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2324 AACAGGCTCTACTATCATATGAGTTGG 2295
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 5
US-09-245-041-3
Sequence 3, Application US/09245041
Patent No. 6274339
GENERAL INFORMATION:
APPLICANT: Moore, K.
APPLICANT: Nagle, D.
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE DIAGNOSIS AND TREATMENT
TITLE OF INVENTION: OF BODY WEIGHT DISORDERS INCLUDING OBESITY
FILE REFERENCE: 7853-136
CURRENT APPLICATION NUMBER: US/09/245,041
CURRENT FILING DATE: 1999-02-05
EARLIER APPLICATION NUMBER: 60/093,630
EARLIER FILING DATE: 1998-07-21
EARLIER APPLICATION NUMBER: 60/104,978
EARLIER FILING DATE: 1998-10-20
NUMBER OF SEQ ID NOS: 131
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 3
LENGTH: 17056
TYPE: DNA
ORGANISM: Mus musculus
US-09-245-041-3

Query Match
Best Local Similarity 1.5%; Score 138.2; DB 3; Length 17056;
Matches 164; Conservative 0; Mismatches 43; Indels 0; Gaps 0;
```



```

RESULT 9
US-07-807-043B-5/C
; Sequence 5, Application US/07807043B
; Patent No. 5342774
; GENERAL INFORMATION:
; APPLICANT: Boon, Thierry, Van den Eynde, Beno t
; TITLE OF INVENTION: Tumor Rejection Antigen Precursors, Tumor
; TITLE OF INVENTION: Rejection Antigens and Uses Thereof
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Felfe & Lynch
; STREET: 805 Third Avenue
; CITY: New York City
; STATE: New York
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage
; COMPUTER: IBM
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/807,043B
; FILING DATE: 19911212
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/764,364
; FILING DATE: 23-SEPTEMBER-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/728,838
; FILING DATE: 9-JULY-1991
; APPLICATION DATA:
; APPLICATION NUMBER: 07/705,702
; FILING DATE: 23-MAY-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Hanson, No. 5342774man D.
; REGISTRATION NUMBER: 30,946
; REFERENCE/DOCKET NUMBER: LUD 253.3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 688-9200
; TELEFAX: (212) 838-3884
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4698 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: singular
; TOPOLOGY: linear
; MOLECULE TYPE: genomic DNA
US-07-807-043B-5

Query Match      1.5%; Score 135.2; DB 1; Length 4698;
Best Local Similarity 86.6%; Pred. No. 9.9e-24;
Matches 149; Conservative 0; Mismatches 23; Indels 0; Gaps 0;

QY 4593 GCATGCTGGCAGACGCTTTAATCCGACACTTGGGAGGAGGAGGAGGAGATTCTG 4652
DB 1706 GCACTCTGCTGTCAGAGCTTAAATCCAGCATTTGGGAGGAGGAGGAGATTCTG 1647
QY 4653 AGTTCAGGCGCAGCGCTGTATAGAGTTCAGGAGGAGGAGGAGGAGGAGGAGGAG 4712
DB 1646 AGTTGAGGCGCAGCGCTGTATAGAGTTCAGGAGGAGGAGGAGGAGGAGGAGGAG 1587
QY 4713 ACCCTGTTTGAAGAAACGAGAAACAAACAAACAAACAAACAAACAAACAAAC 4764
DB 1586 ACCCTGCTCGAAAAACCAAAAAACCAAAAAACCAAAAAACCAAAAAAGGCC 1535

RESULT 10
US-08-299-849B-5/C
; Sequence 5, Application US/08299849B
; Patent No. 5612201
; GENERAL INFORMATION:
; APPLICANT: De Plaen, Etienne; Boon-Falleur, Thierry;
; APPLICANT: Leth, Bernard; Szikora, Jean-Pierre; De Smet, Charles;

```

```

; APPLICANT: Chomez, Patrick
; TITLE OF INVENTION: Isolated Nucleic Acid Molecules Useful In
; TITLE OF INVENTION: Determining Expression Of A Tumor Antigen Precursor
; NUMBER OF SEQUENCES: 48
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Felfe & Lynch
; STREET: 805 Third Avenue
; CITY: New York City
; STATE: New York
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage
; COMPUTER: IBM
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/299,849B
; FILING DATE: 1-SEPTEMBER-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/037,230
; FILING DATE: 26-MARCH-1993
; APPLICATION DATA:
; APPLICATION NUMBER: PCT/US92/04354
; FILING DATE: 22-MAY-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/807,043
; FILING DATE: 12-DECEMBER-1991
; APPLICATION DATA:
; APPLICATION NUMBER: 07/764,364
; FILING DATE: 23-SEPTEMBER-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 9-JULY-1991
; APPLICATION DATA:
; APPLICATION NUMBER: 07/705,702
; FILING DATE: 23-MAY-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Hanson, No. 5612201man D.
; REGISTRATION NUMBER: 30,946
; REFERENCE/DOCKET NUMBER: LUD 5355
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 838-3884
; TELEFAX: (212) 838-3884
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4698 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: genomic DNA
US-08-299-849B-5

Query Match      1.5%; Score 135.2; DB 1; Length 4698;
Best Local Similarity 86.6%; Pred. No. 9.9e-24;
Matches 149; Conservative 0; Mismatches 23; Indels 0; Gaps 0;

QY 4593 GCATGCTGGCAGACGCTTTAATCCGACACTTGGGAGGAGGAGGAGGAGATTCTG 4652
DB 1706 GCACTCTGCTGTCAGAGCTTAAATCCAGCATTTGGGAGGAGGAGGAGGAGATTCTG 1647
QY 4653 AGTTCAGGCGCAGCGCTGTATAGAGTTCAGGAGGAGGAGGAGGAGGAGGAGGAG 4712
DB 1646 AGTTGAGGCGCAGCGCTGTATAGAGTTCAGGAGGAGGAGGAGGAGGAGGAGGAG 1587
QY 4713 ACCCTGTTTGAAGAAACGAGAAACAAACAAACAAACAAACAAACAAACAAAC 4764
DB 1586 ACCCTGCTCGAAAAACCAAAAAACCAAAAAACCAAAAAACCAAAAAAGGCC 1535

RESULT 11
US-08-142-368A-5/C
; Sequence 5, Application US/08142368A

```

Patent No. 5925729

GENERAL INFORMATION:

APPLICANT: Boon-Falleur, Thierry; Van der Bruggen, Thierry;  
APPLICANT: Van den Eynde, Beno t; Van Pel, Aline; De Plaen,  
APPLICANT: Luirquin, Christophe; Chomez, Patrick; Traversari, Catia  
TITLE OF INVENTION: Tumor Rejection Antigen Precursors, Tumor  
TITLE OF INVENTION: Rejection Antigen Precursors, Tumor  
NUMBER OF SEQUENCES: 26

CORRESPONDENCE ADDRESS:  
ADDRESSEE: Felle & Lynch  
STREET: 805 Third Avenue  
CITY: New York City  
STATE: New York  
ZIP: 10022

COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage  
COMPUTER: IBM  
OPERATING SYSTEM: PC-DOS  
SOFTWARE: Wordperfect  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/142,368A  
FILING DATE: 02-MAY-1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US92/04354  
FILING DATE: 22-MAY-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/807,043  
FILING DATE: 12-DECEMBER-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/764,364  
FILING DATE: 23-SEPTEMBER-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/728,838  
APPLICATION NUMBER: 9-JULY-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/705,702  
FILING DATE: 23-MAY-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Hanson, No. 5925729man D.  
REGISTRATION NUMBER: 30,946  
REFERENCE/DOCKET NUMBER: LUD 5253.4-US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 688-9200  
TELEFAX: (212) 838-3884  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4698 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: genomic DNA

US-08-142-368A-5



Search completed: September 13, 2003, 19:53:41  
Job time : 441 secs

```

RESULT 15
US-09-579-197-5/C
: Sequence 5, Application US/09579197
: Patent No. 6552180
: GENERAL INFORMATION:
: APPLICANT: Gaugler, Beatrice; Van den Eynde, Benoît;
: APPLICANT: van der Bruggen, Pierre; Boon-Falleur, Thierry
: TITLE OF INVENTION: Isolated Nucleic Acid Molecules Coding
: TITLE OF INVENTION: For Tumor Rejection Antigen Precursor Mage-3 And Uses Thereof
: NUMBER OF SEQUENCES: 30
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Felle & Lynch
: STREET: 805 Third Avenue
: CITY: New York City
: STATE: New York
: ZIP: 10022
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage
: COMPUTER: IBM
: OPERATING SYSTEM: PC-DOS
: SOFTWARE: Wordperfect
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/579,197
: FILING DATE:
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/037,230
: FILING DATE:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 07/807,043
: FILING DATE: 12-DECEMBER-1991
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 07/764,364
: FILING DATE: 23-SEPTEMBER-1991
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 07/728,838
: FILING DATE: 9-JULY-1991
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 07/705,702
: FILING DATE: 23-MAY-1991
: ATTORNEY/AGENT INFORMATION:
: NAME: Hanson, No. 6552180man D.
: REGISTRATION NUMBER: 30,946
: REFERENCE/DOCKET NUMBER: LUD 5353
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (212) 688-9200
: TELEFAX: (212) 838-3884
: INFORMATION FOR SEQ. ID NO.: 5:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 4698 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: genomic DNA
: US-09-579-197-5

Query Match      1.5%; Score 135.2; DB 4; Length 4698;
Best Local Similarity 86.6%; Pred. No. 9.9e-24;
Matches 149; Conservative 0; Mismatches 23; Indels 0; Gaps 0;

QY      4593 GCAGTGGTGACACGCCCTTTAATCCAGCACTTGGAGGACAGACAGCAGGATTTCTG 4652
DB      1706 GCAGTCTCTGGTCAAGCTTTAATCCAGCATTTGGAGGACAGACAGCAGCATTTCTG 1647
QY      4653 AGTTCAAGGCCAGCCTGGTCTATAGAGTAGTTCAGACAGCAGGCGCTACACAGAA 4712
DB      1646 AGTTTAGGCCAGCCTGGTCTACAGAGTAGTGCCAGACAGCAGGATACAAAGAA 1587
QY      4713 ACCGTGTTTGAACACAGAAAAACAAAAACAAAAACAAAAACAAAAAC 4764
DB      1586 ACCGTGTTCTGAAAAACCAAAAAACAAAAACAAAAACAAAAACAGGCC 1535

```





XX WO200034492-A1.  
 PN 15-JUN-2000.  
 PD 09-DEC-1998; 98WO-EP08009.  
 PE 09-DEC-1998; 98WO-EP08009.  
 XX 09-DEC-1998; 98WO-EP08009.  
 XX (CNRS ) CENT NAT RECH SCI.  
 XX (CURIE-) INST CURIE.  
 PA  
 XX  
 XX Pinto D, Robine S, Jaisser F, Louvard D;  
 DR WPI; 2000-423433/36.  
 XX  
 PT Novel nucleotide sequence derived from mouse villin gene for targeted  
 PT expression of transgenes in immature and differentiated epithelial  
 PT cells of intestine or urogenital tracts -  
 XX  
 XX  
 PS Claim 3; Fig 6; 54pp; English.  
 XX  
 CC The present sequence comprises the murine villin gene regulatory  
 CC region, the first intron and part of the first two exons. It has been  
 CC shown that this region directs the expression of the villin gene in the  
 CC intestine and urogenital tracts, and thus could be used in a fusion  
 CC gene to direct expression of exogenous genes in these areas. This could  
 CC be used, for example, to create a mouse model for colorectal cancer.  
 CC  
 XX Sequence 8995 BP; 2275 A; 2105 C; 2258 G; 2357 T; 0 other;  
 XX

Query Match	100.0%;	Score 8995;	DB 21;	Length 8995;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 8995;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

OY	I	GATCTGGTGACCAAGACACTGTGSTGCCAGCACTGGGAGGTGGAGGAGGCTGA	60
Dd	1	GATCTGGTGACCAAGGACACTGTGTCCTCCACACTGGGAGGATGGAGGAGGCTCA	60
OY	61	GAAGTTTAAGGCATCCTTGGTTACATTACCAGGTTTTCAGCCAGCTTCAGCTACATGA	120
Dd	61	GAAGTTTAAGGCATCCTTGGTTACATATACAAGGTTTCAGCCAGCTTCAGCTACATGA	120
OY	121	CCTTTGTTGGTTTGGTTTGTGTTTAAAGAATTAATAATATATACCATAAGAGGTTGG	180
Dd	121	CCTTTGTTGGTTTGGTTTGTGTTTAAAGCAATTAATAATATATACCATAAGAGGTTGG	180
OY	181	CAGTGGTGCGACACCTTTAATTCAGATATTCAGAGGACAGAACGCGCACATCTCTGT	240
Dd	181	CAGTGGTGCGACACCTTTAATTCAGATATTCAGAGGACAGAACGCGCACATCTCTGT	240
OY	241	GAGTTGGAATGACGCTTAGTCTGCAAAAGCTAGTTCAGAGATGGCAAGGCTACACAGAGA	300
Dd	241	GAGTTGGAATGACGCTTAGTCTGCAAAAGCTAGTTCAGAGATGGCAAGGCTACACAGAGA	300
OY	301	AACCTGTGCTCAATAAACCAAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAG	360
Dd	301	AACCTGTGCTCAATAAACCAAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAG	360
OY	361	GTCCATTAGGATGAGACCATCCTATTAAGATGATTCCTGATCCACGAGTAGGCTAATGTCA	420
Dd	361	GTCCATTAGGATGAGACCATCCTATTAAGATGATTCCTGATCCACGAGTAGGCTAATGTCA	420
OY	421	TGGGAAAAGGGATGAGACTGTCTAGATTAAAGATGCTGAGGCGATCCTATTCTCAA	480
Dd	421	TGGGAAAAGGGATGAGACTGTCTAGATTAAAGATGCTGAGGCGATCCTATTCTCAA	480
OY	481	TTTGATTCATATGAAAAAGGCTGATTAAGGCCCAAGAGAGTAGGAACTGGAGCTCTGGACT	540
Dd	481	TTTGATTCATATGAAAAAGGCTGATTAAGGCCCAAGAGAGTAGGAACTGGAGCTCTGGACT	540
OY	541	GAAGACGTGACGGCCTTATTAACACTGCGACTTATTAACACTTATTAACACTGCGACAG	600
Dd	541	GAAGACGTGACGGCCTTATTAACACTGCGACTTATTAACACTTATTAACACTGCGACAG	600

Db	541	GAAGACGTGACGGCCCTTATTAACACTGCGACTTATTAACACTTATTAACACTGTGCACAGG	600
QY	601	CGTTCAGGTTTGAAGATACCTTTTAAACACAGAAACAGAAAGTGCTGCTCGTCCAGC	660
Db	601	CGTTGAGGTTTGAAGATACCTTTCAAAACACAGAAAGAAAGTGCTGCTCGTCCAGC	660
QY	661	GTAGGAGCACTGGCTGCGAAGAAGTGTATTTAGTGAAAGCTACCTTACACATATCTTT	720
Db	661	GTAGGAGCACTGGCTGCGAAGAAGTGTATTTAGTGAAAGCTACCTTACCAATATCTTT	720
QY	721	GCACCTTATCACATACAGGTGCACAAATGTCATACCTCCCTAGTCACAGATGSGCTGTACA	780
Db	721	GCACCTTATCACATACAGGTGCACAAATGTCATACCTCCCTAGTCACAGATGSGCTGTACA	780
QY	781	CTCGTTTCTGCTTTCCCATCTGTTGTGACATTTGTGCAGAACACAGAAATTAGAAATGTGGT	840
Db	781	CTCGTTTCTGCTTTCCCATCTGTTGTGACATTTGTGCAGAACACAGAAATTAGAAATGTGGT	840
QY	841	ATTTATTTGTGTGCTGAGGACACCATCCAGGCTTTTACATTTTCAGGACATGTTTAC	900
Db	841	ATTTATTTGTGTGCTGAGGACACCATCCAGGCTTTTACATTTTCAGGACATGTTTAC	900
QY	901	TACAGGGGCTACTCTCCACAGGGTTTGAACCATTTGTTTATATTTATATTTACTATTTTGCT	960
Db	901	TACAGGGGCTACTCTCCACAGGGTTTGAACCATTTGTTTATATTTATATTTACTATTTTGCT	960
QY	961	GCATGAGGTAGGACATGTATACGTATGTATAGAGTTCATGCAATGTGGGCTGCTACCCCTCAA	1020
Db	961	GCATGAGGTAGGACATGTATACGTATGTATAGAGTTCATGCAATGTGGGCTGCTACCCCTCAA	1020
QY	1021	ATCATTTGCAGATCCCCACGACAGTGAAGTACCCGAGCGTTGTAACTTGTATGTGGACTG	1080
Db	1021	ATCATTTGCAGATCCCCACGACAGTGAAGTACCCGAGCGTTGTAACTTGTATGTGGACTG	1080
QY	1081	GGAGCGAAGGTGGGGTCTCTGCAAGAGACACGACGTGGCCCTTAACCAATGGGACACAGCT	1140
Db	1081	GGAGCGAAGGTGGGGTCTCTGCAAGAGACACGACGTGGCCCTTAACCAATGGGACACAGCT	1140
QY	1141	CTAGGCGCTAAGAGTATCTTTAGTTTAAAAATATATATCTCAGCGCGGTGTGTGGGC	1200
Db	1141	CTAGGCGCTAAGAGTATCTTTAGTTTAAAAATATATATCTCAGCGCGGTGTGTGGGC	1200
QY	1201	ACAGGCGCTTATATCCACAGCACTTGAGAGGCTGAGAGTGTAGAAATATACACACAGGCCAG	1260
Db	1201	ACAGGCGCTTATATCCACAGCACTTGAGAGGCTGAGAGTGTAGAAATATACACACAGGCCAG	1260
QY	1261	CTGGGGTGCACAGGCTGGCCCTGTGTTTTTTTTTTTTTTTTTTCTTATGTGCACGTGGTCTTA	1320
Db	1261	CTGGGGTGCACAGGCTGGCCCTGTGTTTTTTTTTTTTTTTTTTCTTATGTGCACGTGGTCTTA	1320
QY	1321	CCCTCGTGTATGTCCGTCGACAGGGTGTGCAGATGCCCTTGAGAGCTGAGTTAAAGACAGTTG	1380
Db	1321	CCCTCGTGTATGTCCGTCGACAGGGTGTGCAGATGCCCTTGAGAGCTGAGTTAAAGACAGTTG	1380
QY	1381	TGATTCAGCTGCCCTTACAGATGCTGCGAAATTTGAACCCAGGTGTCCCTAGAGAGCAGCC	1440
Db	1381	TGATTCAGCTGCCCTTACAGATGCTGCGAAATTTGAACCCAGGTGTCCCTAGAGAGCAGCC	1440
QY	1441	AGTGTCTTAACTTCTGAGCCACCCCTCCACCCCGCTTTAAGAGACTCTTAACCTTTTG	1500
Db	1441	AGTGTCTTAACTTCTGAGCCACCCCTCCACCCCGCTTTAAGAGACTCTTAACCTTTTG	1500
QY	1501	TGTATATGTGGAGAACGTAGTGTGATCTGTGCATTTACCAAGTGTGTGCTGGCTGTAGCATCA	1560
Db	1501	TGTATATGTGGAGAACGTAGTGTGATCTGTGCATTTACCAAGTGTGTGCTGGCTGTAGCATCA	1560
QY	1561	CTGAGCCCGTACCCACACGACTAGTGTGATACAGTTTAAAGGCAAAACATTAAACATGACA	1620
Db	1561	CTGAGCCCGTACCCACACGACTAGTGTGATACAGTTTAAAGGCAAAACATTAAACATGACA	1620
QY	1621	ATAGTTGGATAGAGCTTTGATATATATATGTCTGTGACTATTTGTTAGCGTGAACCTTGTCTCC	1680
Db	1621	ATAGTTGGATAGAGCTTTGATATATATATGTCTGTGACTATTTGTTAGCGTGAACCTTGTCTCC	1680

OY	1661	TTAGCATGTGCTGTGAGACAGTATGAAAAATGAGACCTTGAGTCTAGTCTCGGAACCCACA	1740
Db	1681	TTAGCATGTGCTGTGAGAGAGATGAAAAATGAAACCTTGAGTCTAGTCTCGGAACCCACA	1740
OY	1741	GAGGACGGCGAAGAACCCACCTCCGTAAGTGTCTCTGAGGCTGTACATACAACTTGCAT	1800
Db	1741	GAGGACGGCGAAGAACCCACCTCCTAAGATGTTCTCTGAGGCTGTACATACAACTTGCAT	1800
OY	1801	AATAGTTTCAATGATTAATTAATTAATTAATTTAGTAATCTTTTAAAGGTATATGTTGGGAGGA	1860
Db	1801	AATAGTTTCAATGATTAATTAATTAATTAATTTAGTAATTTCTTTTAAAGGTATATGTTGGGAGGA	1860
OY	1861	GAGATGGCTACGTTCTCCAGAGACCTTGCTGCTCTTGCGAGGACCTTAGATTTCAGTTCC	1920
Db	1861	GAGATGGCTACGTTCTCCAGAGACCTTGCTGCTCTTGCGAGGACCTTAGATTTCAGTTCC	1920
OY	1921	AGGACTCATATGGTGGCTCACAGGCACTGTGTAATCCAGTCCGAGGGTTCCACACCT	1980
Db	1921	AGGACTCATATGGTGGCTCACAGGCACTGTGTAATCCAGTCCGAGGGTTCCACACCT	1980
OY	1981	CTTGTGGCTTCACAGGACCCACATATAGTATACAGACATATCATGTCAGGCAAAACAC	2040
Db	1981	CTTGTGGCTTCACAGGACCCACATATAGTATATACAGACATATCATGAGCAAAACAC	2040
OY	2041	CATACACACATTAATTAATTAAGAACTTTAAAGGTGCATGTGTTGCTAAATTTGCT	2100
Db	2041	CATACACACATTAATTAATTAAGAACTTTAAAGGTGCATGTGTTGCTAAATTTGCT	2100
OY	2101	TACACATGCTGATTTGAAGACATGTACAAACGACACCTGAAAGAGGATCTGGGCTCGAG	2160
Db	2101	TACACATGCTGATTTGAAGACATGTACAAACGACACCTGAAAGAGGATCTGGGCTCGAG	2160
OY	2161	AGATGGCTCAGCGGTTTAAAGACACTGACTGCTCTCCGAAGAGAGTCTGAGTTCAAAT	2220
Db	2161	AGATGGCTCAGCGGTTTAAAGACACTGACTGCTCTCCGAAGAGAGTCTGAGTTCAAAT	2220
OY	2221	CCTGACAAACCATGTGGCTCACACACCATCATTAATGAGATCTGACACCTCTTGTGCT	2280
Db	2221	CCTGACAAACCATGTGGCTCACACACCATCATTAATGAGATCTGACACCTCTTGTGCT	2280
OY	2281	GCATCTGAGACAGCTGCGAGACTACAGTGTACTAGTATACATAATTAATTAATCTTTT	2340
Db	2281	GCATCTGAGACAGCTGCGAGACTACAGTGTACTAGTATACATAATTAATTAATCTTTT	2340
OY	2341	TTTTAAAAAATGAAGAGGATCTGAGACACCTCAAAAAGAGATTTATGACAGTCACTCAG	2400
Db	2341	TTTTAAAAAATGAAGAGGATCTGAGACACCTCAAAAAGAGATTTATGACAGTCACTCAG	2400
OY	2401	GGTATATATCTATCTCTGGAGTTTTCCTTTCCTGGCTTGCGAAGTGGTGGACACAG	2460
Db	2401	GGTATATATCTATCTCTGGAGTTTTCCTTTCCTGGCTTGCGAAGTGGTGGACACAG	2460
OY	2461	CCCCCTTTTCATTCACAAGAAAGGGGTGCATTTATTTCTGAAACAAAACAGCACTGGAGT	2520
Db	2461	CCCCCTTTTCATTCACAAGAAAGGGGTGCATTTATTTCTGAAACAAAACAGCACTGGAGT	2520
OY	2521	ATGTTTAACTGTCTTGGCTATGAGCAGCGCAGCGCGCGCGGCGACACACACACAC	2580
Db	2521	ATGTTTAACTGTCTTGGCTATGAGCAGCGCAGCGCGCGCGCGGCGACACACACACAC	2580
OY	2581	ACACACACACACACACACACACACACACATTCAGTCTCCAGAGCTCTTGGGAAGTCA	2640
Db	2581	ACACACACACACACACACACACACACACATTCAGTCTCCAGAGCTCTTGGGAAGTCA	2640
OY	2641	AGAAAGAGGCTCCCTCAAAACAGATCTTCAATCTTCCCTCCTTAAAGAGACACAGATTC	2700
Db	2641	AGAAAGAGGCTCCCTCAAAACAGATCTTCAATCTTCCCTCCTTAAAGAGACACAGATTC	2700
OY	2701	AAGGTGCAGAAATCTTACAGGGGGCGAGGCGAGGGAAGGGAAGCAAGGCATGTTTCC	2760
Db	2701	AAGGTGCAGAAATCTTACAGGGGGCGAGGCGAGGGAAGGGAAGCAAGGCATGTTTCC	2760

QY	2761	AGAGACCTACAGACAGAGGGCAGACAGAGATGCCAGTCCAGGGCAGGGAGGTGGAGG	2820
Db	2761	AGAGACCTACAGACAGAGGGCAGACAGAGATGCCAGTCCAGGGCAGGGAGGTGGAGG	2820
QY	2821	CCCTTGTTCCGAGGAGGAAGGCGAGGGGAGAAACAGGGTTCAAAGGCACAGGTTTATGACA	2880
Db	2821	CCCTTGTTCCGAGGAGGAAGGCGAGGGGAGAAACAGGGTTCAAAGGCACAGGTTTATGACA	2880
QY	2881	GCTCATAAAAGTGAAGTCGTGGCTCACTCAGAAAAGGAGGAAGAAAGGAAAGGCCCTTGT	2940
Db	2881	GCTCATAAAAGTGAAGTCGTGGCTCACTCAGAAAAGGAGGAAGAAAGGAAAGGCCCTTGT	2940
QY	2941	GCCCACTGAGCGAGGGTATGCTGAGTAGAGAGATCTGCAGGGGTGCCAGAGGCCAC	3000
Db	2941	GCCCACTGAGCGAGGGTATGCTGAGTAGAGAGATCTGCAGGGGTGCCAGAGGCCAC	3000
QY	3001	CTGTGTGCCCAAAGGAAACCCCAATGTGAACCTCGGGCTTGGGGTGCAGATCCAGTA	3060
Db	3001	CTGTGTGCCCAAAGGAAACCCCAATGTGAACCTCGGGCTTGGGGTGCAGATCCAGTA	3060
QY	3061	CAAGACCCCAAGAGTCTACTCTCAATCCCATCAATGCCCCCTCGCCCCACACCCCA	3120
Db	3061	CAAGACCCCAAGAGTCTACTCTCAATCCCATCAATGCCCCCTCGCCCCACACCCCA	3120
QY	3121	CCCCGAGTCCCGTCCACTTCTCTAGAGGCTGGAGGGTGGCCAGCCCTGTGGGGTTGC	3180
Db	3121	CCCCGAGTCCCGTCCACTTCTCTAGAGGCTGGAGGGTGGCCAGCCCTGTGGGGTTGC	3180
QY	3181	CTACCTGAGGTAGAGGCCAGGTCTTACCCGGAATGTGACCCCATCCCTGAAGCTGCAGA	3240
Db	3181	CTACCTGAGGTAGAGGCCAGGTCTTACCCGGAATGTGACCCCATCCCTGAAGCTGCAGA	3240
QY	3241	GCCAAAGGCGGGGACACAGGCGAGCTCAGGGCTGTGTGCTGGGCTCTAGTGTCC	3300
Db	3241	GCCAAAGGCGGGGACACAGGCGAGCTCAGGGCTGTGTGCTGGGCTCTAGTGTCC	3300
QY	3301	AGGACCTGGGCACCTTCTCCACCCGCCCATCATTTCTCTGTGGGCCCTATCTTCC	3360
Db	3301	AGGACCTGGGCACCTTCTCCACCCGCCCATCATTTCTCTGTGGGCCCTATCTTCC	3360
QY	3361	CTTATATGTTAAGGAATGCTCGGGGGGGGGGTGTGGTGAAGGAATGAAGTTCGTGG	3420
Db	3361	CTTATATGTTAAGGAATGCTCGGGGGGGGGGTGTGGTGAAGGAATGAAGTTCGTGG	3420
QY	3421	GTCTCTCAGACAGCTTGCCACAATCTCTAAGATCTCCAGAGGTGGTGCCTCTTC	3480
Db	3421	GTCTCTCAGACAGCTTGCCACAATCTCTAAGATCTCCAGAGGTGGTGCCTCTTC	3480
QY	3481	CAGACAGGTAAGGCAATTTGGTGGGACACATGTGTACACAGGTGTGTGAAGGGACAG	3540
Db	3481	CAGACAGGTAAGGCAATTTGGTGGGACACATGTGTACACAGGTGTGTGAAGGGACAG	3540
QY	3541	GGTCTGTGCTTCTCTGTGGCAAGCTGTGCTTCTGTACCACTTGTGTATAGTTTGGGG	3600
Db	3541	GGTCTGTGCTTCTCTGTGGCAAGCTGTGCTTCTGTACCACTTGTGTATAGTTTGGGG	3600
QY	3601	TGAGGTAAAGTGTCTGAACCTCTGAAGAAGCAAGAGCCAGCAGGCTGTCTTGGGCT	3660
Db	3601	TGAGGTAAAGTGTCTGAACCTCTGAAGAAGCAAGAGCCAGCAGGCTGTCTTGGGCT	3660
QY	3661	TCAATGAAGAAGTTACAGACAGCCCTTCTGTAAAGTCAACCTTGCCTTCACTGTGTAG	3720
Db	3661	TCAATGAAGAAGTTACAGACAGCCCTTCTGTAAAGTCAACCTTGCCTTCACTGTGTAG	3720
QY	3721	ATTCCCTGGGACCAAGTGGCTCTGGGACACATTTCTCAATTAATAACAGACAGT	3780
Db	3721	ATTCCCTGGGACCAAGTGGCTCTGGGACACATTTCTCAATTAATAACAGACAGT	3780
QY	3781	CCTGAGATTGAGACCTCGCTGCTGATTTACTACTTCTCTGGCTGCTCATTTCTGTGT	3840
Db	3781	CCTGAGATTGAGACCTCGCTGCTGATTTACTACTTCTCTGGCTGCTCATTTCTGTGT	3840
QY	3841	TCAATGCTTACACATCTGAATGTGTTCTTTGTGTCAACATTCCTCGTACACTCTGGGA	3900

Db 3841 TCATGCTTACACATCTGAAATGTTTCTTGTGTGACCACTTCCCTGACACTCTGGGA 3900  
 QY 3901 GGTGCTATCCTTGGACATGTATCTGGATGTAACTGCAGCCAGCAGAGAGGGG 3960  
 Db 3901 GGTGCTATCCTTGGACATGTATCTGGATGTAACTGCAGCCAGCAGAGAGGGG 3960  
 QY 3961 AGATCAGAGAGCTGTGTCTTCTAGGCTTATTTAGGCTTGACATCACTTCTTCTAGAAAT 4020  
 Db 3961 AGATCAGAGAGCTGTGTCTTCTAGGCTTATTTAGGCTTGACATCACTTCTTCTAGAAAT 4020  
 QY 4021 GGCCCTCCATTTTGGTTTACCATGATCTATTTTATACAGCTGGCGCTAAAGCCA 4080  
 Db 4021 GGCCCTCCATTTTGGTTTACCATGATCTATTTTATACAGAGTGGCGCTAAAGCCA 4080  
 QY 4081 AACCTGCCAGAAAGTTGGGACTCTACAGACCAAGGTTATCTGCTAGAAATCCCTG 4140  
 Db 4081 AACCTGCCAGAAAGTTGGGACTCTACAGACCAAGGTTATCTGCTAGAAATCCCTG 4140  
 QY 4141 TCATCTGAGGTTGGGAGAAATCTGCTGGGGCTTCCAGGCTTGTAGCAGAGGGT 4200  
 Db 4141 TCATCTGAGGTTGGGAGAAATCTGCTGGGGCTTCCAGGCTTGTAGCAGAGGGT 4200  
 QY 4201 ATCTTGTATAGGGAATGACCTAGTCTATGTTTACTATTCCTGCTCAATTAAG 4260  
 Db 4201 ATCTTGTATAGGGAATGACCTAGTCTATGTTTACTATTCCTGCTCAATTAAG 4260  
 QY 4261 CTGGAACTAAACCCACGCGAGCCAGAGATCTCTACAGTTGTACCCCAAGACACA 4320  
 Db 4261 CTGGAACTAAACCCACGCGAGCCAGAGATCTCTACAGTTGTACCCCAAGACACA 4320  
 QY 4321 AGCACTAGATATGCAAGATAGTACTGGGAGAAAGAACTTAACCCCAAG 4380  
 Db 4321 AGCACTAGATATGCAAGATAGTACTGGGAGAAAGAACTTAACCCCAAG 4380  
 QY 4381 AGCACTAGATATGCAAGATAGTACTGGGAGAAAGAACTTAACCCCAAG 4440  
 Db 4381 AGCACTAGATATGCAAGATAGTACTGGGAGAAAGAACTTAACCCCAAG 4440  
 QY 4441 TGAGTTGATAGTACAAAGATGATGATGTTCAATGCGAGATGATGCTATGAGCTG 4500  
 Db 4441 TGAGTTGATAGTACAAAGATGATGATGTTCAATGCGAGATGATGCTATGAGCTG 4500  
 QY 4501 TTGGGAGGCTGAAGCAGAGAGATTTCTATATGTTTGAAGCCAGCTGATAGAGCA 4560  
 Db 4501 TTGGGAGGCTGAAGCAGAGAGATTTCTATATGTTTGAAGCCAGCTGATAGAGCA 4560  
 QY 4561 GACTTGTCTTTAAGAAAAAATGAAAGCCAGCAGAGTGGGACACGCTTAAATCCA 4620  
 Db 4561 GACTTGTCTTTAAGAAAAAATGAAAGCCAGCAGAGTGGGACACGCTTAAATCCA 4620  
 QY 4621 GCACCTTGGAGGAGAGAGCAGGAGATTTCTGAGTTCAAGGCCAGCTGCTATAGAGT 4680  
 Db 4621 GCACCTTGGAGGAGAGAGCAGGAGATTTCTGAGTTCAAGGCCAGCTGCTATAGAGT 4680  
 QY 4681 GACTTCCAGAGCAGCCAGGCTACACAGAGAAACCTGTTTGAAGAAACAGAGAAACAA 4740  
 Db 4681 GACTTCCAGAGCAGCCAGGCTACACAGAGAAACCTGTTTGAAGAAACAGAGAAACAA 4740  
 QY 4741 AACAAAAACAACAAAAACAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC 4800  
 Db 4741 AACAAAAACAACAAAAACAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC 4800  
 QY 4801 TAGGCTGTCTGTAGTGTAGTGTAGTGTGGGACTTCAAGCTTATATATATAGGCC 4860  
 Db 4801 TAGGCTGTCTGTAGTGTAGTGTAGTGTGGGACTTCAAGCTTATATATATAGGCC 4860  
 QY 4861 TTTTATACCTGTACAGACGAGAAAGTTTCTAGTCTGGACACAGTGGACCTTGAGA 4920  
 Db 4861 TTTTATACCTGTGTACAGACGAGAAAGTTTCTAGTCTGGACACAGTGGACCTTGAGA 4920  
 QY 4921 AACTACTCTTGGCAACCAAAATTTCTGGGAAGGCTTCTGAGAGAAAGTGTCCCGAT 4980  
 Db 4921 AACTACTCTTGGCAACCAAAATTTCTGGGAAGGCTTCTGAGAGAAAGTGTCCCGAT 4980

Db 4921 AACTACTCTTGGCAACCAAAATTTCTGGGAAGGCTTCTGAGAGAAAGTGTCCCGAT 4980  
 QY 4981 CAGACTACTGTCTAGAGGACAGAGAGAGGTTGGAAGATTTGGTGGACAGAGATT 5040  
 Db 4981 CAGACTACTGTCTAGAGGACAGAGAGAGGTTGGAAGATTTGGTGGACAGAGATT 5040  
 QY 5041 GGAACAGAGAGAGAGAGGAGGAGGATCCAGATTTCTGAACATGTAGCTGACTTTGGT 5100  
 Db 5041 GGAACAGAGAGAGAGAGGAGGAGGATCCAGATTTCTGAACATGTAGCTGACTTTGGT 5100  
 QY 5101 TCTGTGGGTGACAGAGTGTCCCGAGGATGGGCTGTAGAAAGGGGACAGGGGTGAGCC 5160  
 Db 5101 TCTGTGGGTGACAGAGTGTCCCGAGGATGGGCTGTAGAAAGGGGACAGGGGTGAGCC 5160  
 QY 5161 AATGAGTTCAAGTTGAGGAGACATCCAGCCAGGAGCTTGTGCTGGCAACTTAAAGATG 5220  
 Db 5161 AATGAGTTCAAGTTGAGGAGACATCCAGCCAGGAGCTTGTGCTGGCAACTTAAAGATG 5220  
 QY 5221 AGAGCCCTTAACCTTCCCTGAAAGTTTAGGGAGACAGAGAGCTGAGAGATCTTCTA 5280  
 Db 5221 AGAGCCCTTAACCTTCCCTGAAAGTTTAGGGAGACAGAGAGCTGAGAGATCTTCTA 5280  
 QY 5281 GGGTGAAGAGAGGATATCTGCTGACCAACATGGGCTAGAGAGCAGAAAGCAGTTGACAG 5340  
 Db 5281 GGGTGAAGAGAGGATATCTGCTGACCAACATGGGCTAGAGAGCAGAAAGCAGTTGACAG 5340  
 QY 5341 TTACCCCTCAGAACACAGCCATCCCTTGGCTCTAAGAGAGGCTGGGCCCTTCTGTTT 5400  
 Db 5341 TTACCCCTCAGAACACAGCCATCCCTTGGCTCTAAGAGAGGCTGGGCCCTTCTGTTT 5400  
 QY 5401 AAGAACTTACTTTTCTTCAGAGAGAGGACAGAGCTTGTGCCCTCCCTGTTGGTCAA 5460  
 Db 5401 AAGAACTTACTTTTCTTCAGAGAGAGGACAGAGCTTGTGCCCTCCCTGTTGGTCAA 5460  
 QY 5461 TAAACACCCCTGTGTGTAACATTAGTTTACTTCACTGAGTTTGTGCTCAGAGACAGTCA 5520  
 Db 5461 TAAACACCCCTGTGTGTAACATTAGTTTACTTCACTGAGTTTGTGCTCAGAGACAGTCA 5520  
 QY 5521 TCTGTGATAGACTGTCTCTTAACCTACCAAGATATGGCCCAACTTCTCAGAGAGAG 5580  
 Db 5521 TCTGTGATAGACTGTCTCTTAACCTACCAAGATATGGCCCAACTTCTCAGAGAGAG 5580  
 QY 5581 TGCAGAGAGAGGCTTGAAGAAAGGTTAAAGTAAACAAATATGGCCAGATTAACAAACAA 5640  
 Db 5581 TGCAGAGAGAGGCTTGAAGAAAGGTTAAAGTAAACAAATATGGCCAGATTAACAAACAA 5640  
 QY 5641 ACTACTATCTTTGTACCAAAATTTGTTTGTGTAACAGAGAGGAGGTTGTAGTGTAT 5700  
 Db 5641 ACTACTATCTTTGTACCAAAATTTGTTTGTGTAACAGAGAGGAGGTTGTAGTGTAT 5700  
 QY 5701 GT 5760  
 Db 5701 GT 5760  
 QY 5761 CTGGGGGACTTTTCAATGCTAAAGAAATATCTGATATTTGGGCCCATGCCAAGGGTAT 5820  
 Db 5761 CTGGGGGACTTTTCAATGCTAAAGAAATATCTGATATTTGGGCCCATGCCAAGGGTAT 5820  
 QY 5821 TGGGAGAGAGCAGGCTTCTGCAAAACAGTAACTGCCCAAGATGATTTGGTGGCTGAA 5880  
 Db 5821 TGGGAGAGAGCAGGCTTCTGCAAAACAGTAACTGCCCAAGATGATTTGGTGGCTGAA 5880  
 QY 5881 TCACCAAGGGGAGGCTGATCAGAGTGAGCAGAAATCACAAGATTAAGCCACCTGTGGG 5940  
 Db 5881 TCACCAAGGGGAGGCTGATCAGAGTGAGCAGAAATCACAAGATTAAGCCACCTGTGGG 5940  
 QY 5941 GCTCAGAGAGGAGGATTTACAAAGAGTTAAAGGCCAAGCCATTTATATCCAAACAGAGAC 6000  
 Db 5941 GCTCAGAGAGGAGGATTTACAAAGAGTTAAAGGCCAAGCCATTTATATCCAAACAGAGAC 6000  
 QY 6001 TCAAAATCAAAAGTGCAAGGAGATTTAGCTGGAGAGATGGGGCTGCAGTGTGGAGACCC 6060  
 Db 6001 TCAAAATCAAAAGTGCAAGGAGATTTAGCTGGAGAGATGGGGCTGCAGTGTGGAGACCC 6060

OY	6061	1GACCTTGCACTTATTA	GTAGTCACTAC	TAGGCCAAGGACAGT	CACAGAGGGTACTGSGGTCTTA	6120
Db	6061	1GACCTTGCACTTATTA	GTAGTCACTAC	TAGGCCAAGGACAGT	CACAGAGGGTACTGSGGTCTTA	6120
OY	6121	CTCAGCTTGGAGGACG	GCAGTGGAGAA	TGGGTGGAC	CTCCATCCGATGGAGAGGGCTGAG	6180
Db	6121	CTCAGCTTGGAGGACG	GCAGTGGAGAA	TGGGTGGAC	CTCCATCCGATGGAGAGGGCTGAG	6180
OY	6181	CACGACGATACAGT	CTTCCCTGTGTCTAT	CCGAGGATTCCTGGCCAGTTT	TTTCAAG	6240
Db	6181	CACGACGATACAGT	CTTCCCTGTGTCTAT	CCGAGGATTCCTGGCCAGTTT	TTTCAAG	6240
OY	6241	GACTTAAAGACTCAT	CTCTGCTGGTGA	AAACAATATCCAAAGCCCTTAAGCCCA	ATTTGGTCT	6300
Db	6241	GACTTAAAGACTCAT	CTCTGCTGGTGA	AAACAATATCCAAAGCCCTTAAGCCCA	ATTTGGTCT	6300
OY	6301	AATTAAATCAACA	CCCCCGGGGATGAC	GGCTGTGAGCAGCAGGAGCTTTT	AAAAAGCTC	6360
Db	6301	AATTAAATCAACA	CCCCCGGGGATGAC	GGCTGTGAGCAGCAGGAGCTTTT	AAAAAGCTC	6360
OY	6361	CCAGGTGATTTCTG	ATCCAGCAGCTG	GAACAAACAGCTACAGTTC	CAACGAAAGAGGC	6420
Db	6361	CCAGGTGATTTCTG	ATCCAGCAGCTG	GAACAAACAGCTACAGTTC	CAACGAAAGAGGC	6420
OY	6421	AAAGCTAGGAAAG	CGTTGGATGGGAG	CCCTTCCAGGCACTAGATGGAAGGCTGGT		6480
Db	6421	AAAGCTAGGAAAG	CGTTGGATGGGAG	CCCTTCCAGGCACTAGATGGAAGGCTGGT		6480
OY	6481	AGCAGTGTGGCAG	CTTCTCTGCTGTCAT	ATAGTATCCATCCACTCATCATCAT		6540
Db	6481	AGCAGTGTGGCAG	CTTCTCTGCTGTCAT	ATAGTATCCATCCACTCATCATCAT		6540
OY	6541	ACACCCACCCATTC	ATATGACACCAATCCCTTCATCAT	ATCATCTATCCAGCTACCCAC		6600
Db	6541	ACACCCACCCATTC	ATATGACACCAATCCCTTCATCAT	ATCATCTATCCAGCTACCCAC		6600
OY	6601	CCAGGCATCATTC	CAAAACCTTCCTTCTCTCTTCTTCTTCTT	CTTTCCTTCACTCAT		6660
Db	6601	CCAGGCATCATTC	CAAAACCTTCCTTCTCTCTTCTTCTTCTT	CTTTCCTTCACTCAT		6660
OY	6661	CATTATTCACA	CAGAACTGCTAT	TGTACTAAATGTGGAGATTA	TAATTTTGA	6720
Db	6661	CATTATTCACA	CAGAACTGCTAT	TGTACTAAATGTGGAGATTA	TAATTTTGA	6720
OY	6721	AGCTCTGTGATTA	CTACTATGTGTCAT	GTATGTGAGACAGTACAT	TACACGACACAGCTG	6780
Db	6721	AGCTCTGTGATTA	CTACTATGTGTCAT	GTATGTGAGACAGTACAT	TACACGACACAGCTG	6780
OY	6781	TGGCAATCGA	AAAGGTTTGGGTGTGTTTCTCTCCACCGCT	GTGGGTTTG66GA		6840
Db	6781	TGGCAATCGA	AAAGGTTTGGGTGTGTTTCTCTCTCCACCGCT	GTGGGTTTG66GA		6840
OY	6841	TTGAAC	TCAATTAATTCGGGGGTGGTGGGCAAGT	GTCTTACCAACGACGACATTTGCTACA		6900
Db	6841	TTGAAC	TCAATTAATTCGGGGGTGGTGGGCAAGT	GTCTTACCAACGACGACATTTGCTACA		6900
OY	6901	CATCATTA	TTATTAAGAAACATCTT	ATGTAGTCCAGGCTGGCC	TAAGCTTGCTATGTCG	6960
Db	6901	CATCATTA	TTATTAAGAAACATCTT	ATGTAGTCCAGGCTGGCC	TAAGCTTGCTATGTCG	6960
OY	6961	CCAGGATAGAC	CTTTAACTCCTGCTCTTCCAGCCCTCA	CCCGAATGCTAGTTTACAGGT		7020
Db	6961	CCAGGATAGAC	CTTTAACTCCTGCTCTTCCAGCCCTCA	CCCGAATGCTAGTTTACAGGT		7020
OY	7021	GTTCAACTG	GTGAATGCTTTAATCC	AGCACTGTGGGGGGGGGGGGAGGACGATC		7080
Db	7021	GTTCAACTG	GTGAATGCTTTAATCC	AGCACTGTGGGGGGGGGGGGAGGACGATC		7080
OY	7081	CCTGAGTTG	GAGCCAGTTTGGTCTAC	AGAGTTCAGAGATCTG	GGGGCTATACAGGAA	7140
Db	7081	CCTGAGTTG	GAGCCAGTTTGGTCTAC	AGAGTTCAGAGATCTG	GGGGCTATACAGGAA	7140

OY	7141	ACCCATTCGCCAAACAACCAACCAACAAAAAATATTCTGTGCATAATATCAGAGA	7200
Dd	7141	ACCCTATTCGCCAAACAACCAACCAACAAAAAATTTCTGTGCAATATATCACAGA	7200
OY	7201	TTAGGGATTAATTAGAGGTAGTGAAGGGGTGTGAGGGGAGTCAATGCTTCTTTGATT	7260
Dd	7201	TTAGGGGTTATTAGTAGGTAAGTGAGGGGTGTGAGGGGAGTCAATGCTTCTTTGATT	7260
OY	7261	ATAATAGTAAAGTACTCATCAAGATGACATTATCTATCTATCTATCTATCTATCTATCT	7320
Dd	7261	ATAATAGTAAAGTACTCATCAAGATGACATTATCTATCTATCTATCTATCTATCTATCT	7320
OY	7321	TATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCT	7380
Dd	7321	TATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCT	7380
OY	7381	CTGCTTTGACCTCAATAGCTCCCTATTTCTGGGTCAACTCTGACCCTAGTGGGTTT	7440
Dd	7381	CTGCTTTGACCTCAATAGCTCCCTATTTCTGGGTCAACTCTGACCCTAGTGGGTTT	7440
OY	7441	ACCAACACCCAGACATTATTTATTTTGGTTTTATTTTATTAATCTAGSAGCTCAGGGT	7500
Dd	7441	ACCAACACCCAGACATTATTTATTTTGGTTTTATTTTATTAATCTAGSAGCTCAGGGT	7500
OY	7501	GGACTCAGAGGCTTTGTGCATCTAGACAAGCTCTGTGCCACAGCTGCAGCTCAGTCC	7560
Dd	7501	GGACTCAGAGGCTTTGTGCATCTAGACAAGCTCTGTGCCACAGCTGCAGCTCAGTCC	7560
OY	7561	CCATTTTTGTAGGTGACTGTGTACAGTGTGTCAATTCGAGGGCTATGATGCTCTCTC	7620
Dd	7561	CCATTTTTGTAGGTGACTGTGTACAGTGTGTCAATTCGAGGGCTATGATGCTCTCTC	7620
OY	7621	CACCTCCAGATTCAGACACTTTCTGTCATCCAGTGGGGGGGCACTCTGTCTACCA	7680
Dd	7621	CACCTCCAGATTCAGACACTTTCTGTGTCATCCAGTGGGGGGGCACTCTGTCTACCA	7680
OY	7681	GTGCCCTGTCCCTGTCTCTAGACCTACATATTTGCTGTCTGAACAGTTCTATTAATG	7740
Dd	7681	GTGCCCTGTCCCTGTCTCTAGACCTACATATTTGCTGTCTGAACAGTTCTATTAATG	7740
OY	7741	GGATGCGTTCCTGTGTATCTTTATATGGCTGGCCCCCTTATCTTAGACAGTTTGTGTG	7800
Dd	7741	GGATGCGTTCCTGTGTATCTTTATATGGCTGGCCCCCTTATCTTAGACAGTTTGTGTG	7800
OY	7801	GGCCTATGTCTAGCTGCTATFACTATCTTATCATCATCATCTATGAGCTTAATAGTTCCT	7860
Dd	7801	GGCCTATGTCTAGCTGCTATFACTATCTTATCATCATCATCTTATGAGCTTAATATGIST	7860
OY	7861	TGTGTGATTAACCACTTCTGTTCATTTACTGATGGAAATTTGTGGCCCCACCCCAC	7920
Dd	7861	TGTGTGATTAACCACTTCTGTTCATTTACTGATGGAAATTTGTGGCCCCACCCCAC	7920
OY	7921	CCTTTTTTTTTTATTTGAGACAAGGCTTCTGTGTAATCTTCCAAATCTGTGGCTGT	7980
Dd	7921	CCTTTTTTTTTTATTTGAGACAAGGCTTCTGTGTAATCTTCCAAATCTGTGGCTGT	7980
OY	7981	GAGCTCACTGTGTAGACAGGCTGTGAGAGGCTGCTCTTCCACTTTTGACACTCCCTGTAA	8040
Dd	7981	GAGCTCACTGTGTAGACAGGCTGTGAGAGGCTGCTCTTCCACTTTTGACACTCCCTGTAA	8040
OY	8041	AGAGTAGCCATGAACCTTCCAAGACAAATTTTCTGTTTGGTTTGTATTTACATTTGTGT	8100
Dd	8041	AGAGTAGCCATGAACCTTCCAAGACAAATTTCTGTTTGGTTTGTATTTACATTTGTGT	8100
OY	8101	TGTATGCGGTATATGTGCATGTTTGTGTCTTCAAGGTGTCTACATGTGTACTGTGTG	8160
Dd	8101	TGTATGCGGTATATGTGCATGTTTGTGTCTTCAAGGTGTCTACATGTGTACTGTGTG	8160
OY	8161	TGGGACAGAGAACAAACCGATGTGCCAATTCCTCGATAGTACAGCATTTGTATATATGA	8220
Dd	8161	TGGGACAGAGAACAAACCGATGTGCCAATTCCTCGATAGTACAGCATTTGTATATATGA	8220
OY	8221	TGTATTATATGTTATTTATTTAGTGTGCCCAAGTATGACAGTATTTCTTGGAGTTTTCAC	8280





QY 421 TGGGGAAGGAGTGGAGCTGCTAGATTAAAAAGTGTGAGGGAGTCCATTCTCAA 480  
| | | | |  
Db 421 TGGGGAAGGAGTGGAGCTGCTAGATTAAAAAGTGTGAGGGAGTCCATTCTCAA 480  
QY 481 TTTGATTCCATATGAAAAGGCTGATAGAGCCCAAGAGAAATGGAATGGAGCTTGAGCT 540  
| | | | |  
Db 481 TTTGATTCCATATGAAAAGGCTGATAGAGCCCAAGAGAAATGGAATGGAGCTTGAGCT 540  
QY 541 GAAGAGCTGAGAGGCTTTAATACATGGAGCTTATTAACCTTATTAACAGTGAACAGGACAG 600  
| | | | |  
Db 541 GAAGAGCTGAGAGGCTTTAATAACTGGAGCTTATTAACCTTATTAACAGTGAACAGGACAG 600  
QY 601 CGTTACAGTTTGAAGATCACTTTCAAACCCACAGAAAGTGTGCTGCTGCTCAGC 660  
| | | | |  
Db 601 CGTTACAGTTTGAAGATCACTTTCAAACCCACAGAAAGTGTGCTGCTGCTCAGC 660  
QY 661 GTAGGAGACACTGGCTGCAGAAAGTGTATTTAGTGAAGAGCTACCTTCACAAATATCTTT 720  
| | | | |  
Db 661 GTAGGAGACACTGGCTGCAGAAAGTGTATTTAGTGAAGAGCTACCTTCACAAATATCTTT 720  
QY 721 GCACCTTATCACATACAGTGTCAAAATGTCTAAGTCCCTGATCCAGATGGCTGTACA 780  
| | | | |  
Db 721 GCACCTTATCACATACAGTGTCAAAATGTCTAAGTCCCTGATCCAGATGGCTGTACA 780  
QY 781 CTGCTTTCTGCTTCCATCTGCTGATGACATTTGTGAGAACAGAAATTTAGAAATGTGGGT 840  
| | | | |  
Db 781 CTGCTTTCTGCTTCCATCTGCTGATGACATTTGTGAGAACAGAAATTTAGAAATGTGGGT 840  
QY 841 ATTTATTTGTGTGCTGAGAGACACCATCCAGGGCTTTTCACATTTACAGGCACATGTGTTAC 900  
| | | | |  
Db 841 ATTTATTTGTGTGCTGAGAGACACCATCCAGGGCTTTTCACATTTACAGGCACATGTGTTAC 900  
QY 901 TTAAGTGGCTACTTCTCCAAAGGTTTGAACCAATTTGTTTATTTACTTATTTTGTGT 960  
| | | | |  
Db 901 TTAAGTGGCTACTTCTCCAAAGGTTTGAACCAATTTGTTTATTTACTTATTTTGTGT 960  
QY 961 GCATGAGGTAGGATGTATACGTATGTATAGAGTATGATGATGATGATGATGATGATGATGAT 1020  
| | | | |  
Db 961 GCATGAGGTAGGATGTATACGTATGTATAGAGTATGATGATGATGATGATGATGATGATGAT 1020  
QY 1021 ATCATTTGACAGATCCCCAGCAAGTGAAGTACCGAGGCTTTGTAAGTTGTTATGTGGACATG 1080  
| | | | |  
Db 1021 ATCATTTGACAGATCCCCAGCAAGTGAAGTACCGAGGCTTTGTAAGTTGTTATGTGGACATG 1080  
QY 1081 GGAGGCAAGGCTGGGCTTCTGCTCAAGAGAGCAGAGGCTTAAACATGGAGCAGCAGCTT 1140  
| | | | |  
Db 1081 GGAGGCAAGGCTGGGCTTCTGCTCAAGAGAGCAGAGGCTTAAACATGGAGCAGCAGCTT 1140  
QY 1141 CTAGGCTTAAGGTAATCTTTAGTTTTTAAAAATATATATTCYCA6CCGGGTGTGGTGC 1200  
| | | | |  
Db 1141 CTAGGCTTAAGGTAATCTTTAGTTTTTAAAAATATATATTCYCA6CCGGGTGTGGTGC 1200  
QY 1201 ACAGCCCTTTAATCCAGCAGCTTTGAGAGGCTGAGGTGTAGAAATATACACACAGGCCAG 1260  
| | | | |  
Db 1201 ACAGCCCTTTAATCCAGCAGCTTTGAGAGGCTGAGGTGTAGAAATATACACACAGGCCAG 1260  
QY 1261 CTGGGCTCAGAGCTTTGGCCCTGTTTTTTTTTTTTTTTTTTTATGTGACACTGTGTCTTA 1320  
| | | | |  
Db 1261 CTGGGCTCAGAGCTTTGGCCCTGTTTTTTTTTTTTTTTTTTTATGTGACACTGTGTCTTA 1320  
QY 1321 CCTGGGTATGTATCCGTGCAGAGGATGTACATCCCTTGAGAGCTGAGATTAAGAGACAGTTG 1380  
| | | | |  
Db 1321 CCTGGGTATGTATCCGTGCAGAGGATGTACATCCCTTGAGAGCTGAGATTAAGAGACAGTTG 1380  
QY 1381 TGATACAGCTGCGCTTACAGATGTGTGAATTTGAACCCAGGTGTCCCTTAGAGACAGGCC 1440  
| | | | |  
Db 1381 TGATACAGCTGCGCTTACAGATGTGTGAATTTGAACCCAGGTGTCCCTTAGAGACAGGCC 1440  
QY 1441 AGTGTCTTAACTTCTGAGGCCAGCCCTCAACCTGTGCTTTAGAGAGCTTTAAACCTTTTG 1500  
| | | | |  
Db 1441 AGTGTCTTAACTTCTGAGGCCAGCCCTCAACCTGTGCTTTAGAGAGCTTTAAACCTTTTG 1500  
QY 1501 TGTAAATGTGGAACTGAGTGTGATCTTGACATTACCAAGTGTGTGCTGTGAGCATCA 1560  
| | | | |

Db 1501 TGTAAATGTGGAACTGAGTGTGATCTTGACACTTACCAAGTGTGTGCTGTGAGCATCA 1560  
| | | | |  
QY 1561 CTGAGCCCTTACCACACAGCACTATGTGATACAGTTTAAAGGCAACACTTAAACATGACA 1620  
| | | | |  
Db 1561 CTGAGCCCTTACCACACAGCACTATGTGATACAGTTTAAAGGCAACACTTAAACATGACA 1620  
QY 1621 ATAGTTGATAGAGTTTGAATATATAGTCTGAGCTATTTGTTAGCCGAGCTTTGGCTGTCC 1680  
| | | | |  
Db 1621 ATAGTTGATAGAGTTTGAATATATAGTCTGAGCTATTTGTTAGCCGAGCTTTGGCTGTCC 1680  
QY 1681 TTAGCATGTGCTGTGAGAAATAGAAAAATGAAAGTCTAGTCTCTGGAACCCACA 1740  
| | | | |  
Db 1681 TTAGCATGTGCTGTGAGAAATAGAAAAATGAAAGTCTAGTCTCTGGAACCCACA 1740  
QY 1741 GAGGACGGGAGAAACCCACTCTGAAAGTTGTTCTGTGAGCTTACATCAACTTACAT 1800  
| | | | |  
Db 1741 GAGGACGGGAGAAACCCACTCTGAAAGTTGTTCTGTGAGCTTACATCAACTTACAT 1800  
QY 1801 AATAGTTACATGAT 1860  
| | | | |  
Db 1801 AATAGTTACATGAT 1860  
QY 1861 GAGATGGCTCAGCTTCCAGAGACACTGTGCTGCTTTCAGAGAGACCTAGATTCAATGCC 1920  
| | | | |  
Db 1861 GAGATGGCTCAGCTTCCAGAGACACTGTGCTGCTTTCAGAGAGACCTAGATTCAATGCC 1920  
QY 1921 AGGACTCATATGTGTGCTCAGAGCCATCTGTATATTCACATTTCCAGAGGCTTCCACACCT 1980  
| | | | |  
Db 1921 AGGACTCATATGTGTGCTCAGAGCCATCTGTATATTCACATTTCCAGAGGCTTCCACACCT 1980  
QY 1981 CTTTGGCCCTCAGAGGACCAAT 2040  
| | | | |  
Db 1981 CTTTGGCCCTCAGAGGACCAAT 2040  
QY 2041 CATACACACAT 2100  
| | | | |  
Db 2041 CATACACACAT 2100  
QY 2101 TACACATGTGATTTGAAGCATGTACACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 2160  
| | | | |  
Db 2101 TACACATGTGATTTGAAGCATGTACACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 2160  
QY 2161 AGATGGCTCAGGCTTAAAGAGCAGTACAGTCTCTTCCGAAGAGAGGCTCTGAGTTCAAT 2220  
| | | | |  
Db 2161 AGATGGCTCAGGCTTAAAGAGCAGTACAGTCTCTTCCGAAGAGAGGCTCTGAGTTCAAT 2220  
QY 2221 CCTAGCAACCATGTGTGCTCAGAACCATTCATATATATATATATATATATATATATATATAT 2280  
| | | | |  
Db 2221 CCTAGCAACCATGTGTGCTCAGAACCATTCATATATATATATATATATATATATATATATAT 2280  
QY 2281 GCATCTGAAGAGAGCTGAGAGCTACAGTGTACTTATATATATATATATATATATATATATAT 2340  
| | | | |  
Db 2281 GCATCTGAAGAGAGCTGAGAGCTACAGTGTACTTATATATATATATATATATATATATATAT 2340  
QY 2341 TTTAAAAAATATAGAGGAGTGTGAGACACCTCAAAAAGATTTATAGCAGTACAGCAG 2400  
| | | | |  
Db 2341 TTTAAAAAATATAGAGGAGTGTGAGACACCTCAAAAAGATTTATAGCAGTACAGCAG 2400  
QY 2401 GGTGATATCTATCTGAGAGTTTTCCTTCCGCTGTGCTGCAACTGAGTGTGAGACAGC 2460  
| | | | |  
Db 2401 GGTGATATCTATCTGAGAGTTTTCCTTCCGCTGTGCTGCAACTGAGTGTGAGACAGC 2460  
QY 2461 CCCCTTTTCATTCACAAAGAACGGGTCTACATTTATTTGGAACAAACAGCACCTGTCACT 2520  
| | | | |  
Db 2461 CCCCTTTTCATTCACAAAGAACGGGTCTACATTTATTTGGAACAAACAGCACCTGTCACT 2520  
QY 2521 ATGTTTACGTCTTGTGCTACTATGAGCAGGAGCAGCCGAGGAGCAGACACACACACAC 2580  
| | | | |  
Db 2521 ATGTTTACGTCTTGTGCTACTATGAGCAGGAGCAGCCGAGGAGCAGACACACACACACAC 2580  
QY 2581 ACACACACACACACACACACACACACATTCAGTGTGCTGTGAGAGCTTTGGAAAGGTCA 2640  
| | | | |

Db 2581 ACACACACACACACACACACACACACATTCAGTCTCCAGAGCTCTTGGGAAGSTCA 2640  
Qy 2641 AGAAGAGGTCCTCCCTAAACAGATCTTTCATCTTCCCTCTAAAGAGACACAGATTC 2700  
Db 2641 AGAAGAGGTCCTCCCTAAACAGATCTTTCATCTTCCCTCTAAAGAGACACAGATTC 2700  
Qy 2701 AAGGTGGCAAAATCTACAGAGGGGGCAGAGGAGGGGGAAGCAGGSCATGTTTCC 2760  
Db 2701 AAGGTGGCAAAATCTACAGAGGGGGCAGAGGAGGGGGAAGCAGGSCATGTTTCC 2760  
Qy 2761 AGAGACCTACAGAGAGGGCAGCAAGAGCAGATCCCCAGGTCCAGAGGAGGAGTGGAG 2820  
Db 2761 AGAGACCTACAGAGAGGGCAGCAAGAGCAGATCCCCAGGTCCAGAGGAGGAGTGGAG 2820  
Qy 2821 CCTTGTTCGAGAGAGAGGCGGAGAACAGGGTTTCAAGGCAAGGTTTATGGA 2880  
Db 2821 CCTTGTTCGAGAGAGAGGCGGAGAACAGGGTTTCAAGGCAAGGTTTATGGA 2880  
Qy 2881 GCTCATAAAAGTGGAGTCTGTGGCTCAGTCAAGAAAGAGAGAGAGGAAAGGCCCTTGT 2940  
Db 2881 GCTCATAAAAGTGGAGTCTGTGGCTCAGTCAAGAAAGAGAGAGAGGAAAGGCCCTTGT 2940  
Qy 2941 GCCCAGTGAAGAGGTCAGTGTAGTAGAGAGATCTGCAGGGGTCCAGAGGCCAC 3000  
Db 2941 GCCCAGTGAAGAGGTCAGTGTAGTAGAGAGATCTGCAGGGGTCCAGAGGCCAC 3000  
Qy 3001 CTGTCTGTCCCAAGGAAAGCCCAAGTGTGAATCTGTGGCTTGGGTGTGAGTTCCAGCTA 3060  
Db 3001 CTGTCTGTCCCAAGGAAAGCCCAAGTGTGAATCTGTGGCTTGGGTGTGAGTTCCAGCTA 3060  
Qy 3061 CAAGACCCCGAGAGTCTTACTCATCCCATCCAGTCCCTCCGCCCCGCCACACCCCA 3120  
Db 3061 CAAGACCCCGAGAGTCTTACTCATCCCATCCAGTCCCTCCGCCCCGCCACACCCCA 3120  
Qy 3121 CCCCCAGTCCCTGACACTTCTCTAGAGGCTGGAGGTCGACAGCCCTGTGGGGTTC 3180  
Db 3121 CCCCCAGTCCCTGACACTTCTCTAGAGGCTGGAGGTCGACAGCCCTGTGGGGTTC 3180  
Qy 3181 CTACCTGTGAGAGAGCCAGGTCCTAGCCGGAATGACACCCATCCCTGAAGCTGAGA 3240  
Db 3181 CTACCTGTGAGAGAGCCAGGTCCTAGCCGGAATGACACCCATCCCTGAAGCTGAGA 3240  
Qy 3241 GCCAAGGGGGGACACAGGCGACTGAGGCTGTCAAGGCTGTGCTGGGCTCTAGGTTCC 3300  
Db 3241 GCCAAGGGGGGACACAGGCGACTGAGGCTGTGAGGCTGTGCTGGGCTCTAGGTTCC 3300  
Qy 3301 AGGGACCTGGGCACTACTTCCCCACCCCCATCCATTCCTCTGGGGCCATCTTCC 3360  
Db 3301 AGGGACCTGGGCACTACTTCCCCACCCCCATCCATTCCTCTGGGGCCATCTTCC 3360  
Qy 3361 CTATATATGTGAAGAACTCTGTGGGGGGGGGTGTGTGAGGCAAAAGTCTGTTCC 3420  
Db 3361 CTATATATGTGAAGAACTCTGTGGGGGGGGGTGTGTGAGGCAAAAGTCTGTTCC 3420  
Qy 3421 GTCTCTGACAGAGCTTGGCACAACCTCTAGATCTCCAGGTGGTGGCTGCTCTTC 3480  
Db 3421 GTCTCTGACAGAGCTTGGCACAACCTCTAGATCTCCAGGTGGTGGCTGCTCTTC 3480  
Qy 3481 CAGACAGATTAAGCAATTGGGTGGGACACATGTTGACCAACAGTGGTTGAGAGGAG 3540  
Db 3481 CAGACAGATTAAGCAATTGGGTGGGACACATGTTGACCAACAGTGGTTGAGAGGAG 3540  
Qy 3541 GGTCTCTGCTTCTCTCTGGCAGCTGTGCTTGTGAGCAGCTTGGTATATAGTTTGGGG 3600  
Db 3541 GGTCTCTGCTTCTCTCTGGCAGCTGTGCTTGTGAGCAGCTTGGTATATAGTTTGGGG 3600  
Qy 3601 TGAAGTAAAGTGTCTGAAGCTCTGAAGAAAGCAAGAGCAGAGGCTCTTGGGCT 3660  
Db 3601 TGAAGTAAAGTGTCTGAAGCTCTGAAGAAAGCAAGAGCAGAGGCTCTTGGGCT 3660  
Qy 3661 TCAATGAAGAAATTCACAGAGCCCTTTCCTTAAGTCACTTCCCTTATCTGTAG 3720  
Db 3661 TCAATGAAGAAATTCACAGAGCCCTTTCCTTAAGTCACTTCCCTTATCTGTAG 3720

Qy 3721 ATTCCCTGGACCAAGGTGGCTCTCTGGAGCTCAGATTTCTACAAATTAATAACAGACAT 3780  
Db 3721 ATTCCCTGGACCAAGGTGGCTCTCTGGAGCTCAGATTTCTACAAATTAATAACAGACAT 3780  
Qy 3781 CCTGAGACTGGAGCTCCGTGCTGATTTACTACTTCTCTGCTGGCTGCTATTTCTGT 3840  
Db 3781 CCTGAGACTGGAGCTCCGTGCTGATTTACTACTTCTCTGCTGGCTGCTATTTCTGT 3840  
Qy 3841 TCAATGTCTTACACATCTGAATGATGTTCTTGTGTACCAATTCCTCCCTGACACTCTGGGA 3900  
Db 3841 TCAATGTCTTACACATCTGAATGATGTTCTTGTGTACCAATTCCTCCCTGACACTCTGGGA 3900  
Qy 3901 GGTGTATCTTGGCAGCATATATCTGTGGATGTAGCTGCAGCCACCAGAGAGAGGGG 3960  
Db 3901 GGTGTATCTTGGCAGCATATATCTGTGGATGTAGCTGCAGCCACCAGAGAGAGGGG 3960  
Qy 3961 AGAGTCAGAGCTGTGTCTTACAGGCCCTATTAAGGCTGTGACATACCCCTTCTAGAAAT 4020  
Db 3961 AGAGTCAGAGCTGTGTCTTACAGGCCCTATTAAGGCTGTGACATACCCCTTCTAGAAAT 4020  
Qy 4021 GGCCCTCTCATTTTTTGGTTACCATGATCTATTTATATCAGAGTGGGAGTAAGACCA 4080  
Db 4021 GGCCCTCTCATTTTTTGGTTACCATGATCTATTTATATCAGAGTGGGAGTAAGACCA 4080  
Qy 4081 AACCTGCCAAGATTTGGAGCTCAGTCTCAGACCAAGGTAATGTCTCAGAAATCCCTTG 4140  
Db 4081 AACCTGCCAAGATTTGGAGCTCAGTCTCAGACCAAGGTAATGTCTCAGAAATCCCTTG 4140  
Qy 4141 TCACTTGAAGTGGGAGAAATCTGCTGTGGGGCTTCAGGCTCTGGTTTAGCAGAGAGGT 4200  
Db 4141 TCACTTGAAGTGGGAGAAATCTGCTGTGGGGCTTCAGGCTCTGGTTTAGCAGAGAGGT 4200  
Qy 4201 ATCTCTTGTATAGGAGCATGACCTAGTCTATGATGTGTACTACATTTCTCTGCTAGTTAAAG 4260  
Db 4201 ATCTCTTGTATAGGAGCATGACCTAGTCTATGATGTGTACTACATTTCTCTGCTAGTTAAAG 4260  
Qy 4261 CTGGAATTAACCCACAGCGAGCGCCAGATTTCTACAGTTGTACCCCAAGAACCA 4320  
Db 4261 CTGGAATTAACCCACAGCGAGCGCCAGATTTCTACAGTTGTACCCCAAGAACCA 4320  
Qy 4321 AGACAGTAGATATGCAAGATAGTAGTGTGGGGAGGAAGAACTTAAACCCCCCAAG 4380  
Db 4321 AGACAGTAGATATGCAAGATAGTAGTGTGGGGAGGAAGAACTTAAACCCCCCAAG 4380  
Qy 4381 GCCACAGGTTCCGTTCCCTAGTTCAACAATGCCAGTATGAGTCTAGTATGGGCTG 4440  
Db 4381 GCCACAGGTTCCGTTCCCTAGTTCAACAATGCCAGTATGAGTCTAGTATGGGCTG 4440  
Qy 4441 TGAAGTGTAGCTACAAAGCATGAGTATGTTTCATGTGTGTATATCTGAGCAC 4500  
Db 4441 TGAAGTGTAGCTACAAAGCATGAGTATGTTTCATGTGTGTATATCTGAGCAC 4500  
Qy 4501 TTGGAGAGCTGAAGCAGAGGATTTGCTATATGTTTGAAGGCGACCTGATATAGAGCGA 4560  
Db 4501 TTGGAGAGCTGAAGCAGAGGATTTGCTATATGTTTGAAGGCGACCTGATATAGAGCGA 4560  
Qy 4561 GACTTGTCTTAAAGAAAAAATGAAGCCACAGAGTGTGGCAGACGCTTAAATCCA 4620  
Db 4561 GACTTGTCTTAAAGAAAAAATGAAGCCACAGAGTGTGGCAGACGCTTAAATCCA 4620  
Qy 4621 GCACTTGGGAGGAGAAAGCAGGAGATTTCTGAGTTCAAGGCCAGGCTGTCTATAGAGT 4680  
Db 4621 GCACTTGGGAGGAGAAAGCAGGAGATTTCTGAGTTCAAGGCCAGGCTGTCTATAGAGT 4680  
Qy 4681 GAGTTCAGAGACAGGCGCTACAGAGAAACCGTTTGAAGAAACAGAAAAACAA 4740  
Db 4681 GAGTTCAGAGACAGGCGCTACAGAGAAACCGTTTGAAGAAACAGAAAAACAA 4740  
Qy 4741 AACAAAAACAAACAAACAAACCAACCAACCAACCAACCTCATCTCATCTCTC 4800  
Db 4741 AACAAAAACAAACAAACAAACCAACCAACCAACCAACCTCATCTCATCTCTC 4800



```

4801 TAGGCTGTGTCTGTAGTGGTAGAGTTTGGGGACCTTCAGCTATATATATATAATAGGCC 4860
4801 TAGGCTGTGTCTGTAGTGGTAGAGTTTGGGGACCTTCAGCTATATATATAATAGGCC 4860
4861 TTTTATATACGTGTGACAGACGAGAAAGTTTCACTGTGGACACAGTGGGACCTGAGAC 4920
4861 TTTTATATACGTGTGACAGACGAGAAAGTTTCACTGTGGACACAGTGGGACCTGAGAC 4920
4921 AAGTACTCCTTGGCCAGCCAAAATTTGGGAAGGCTTCCTGGAGAAAGTGTGCCGAT 4980
4921 AAGTACTCCTTGGCCAGCCAAAATTTGGGAAGGCTTCCTGGAGAAAGTGTGCCGAT 4980
4981 CAGACTACTGTTCTAGAAAGCAGAAAGAGGTTGGAAGAAATGTTGGTGACAGACAGTT 5040
4981 CAGACTACTGTTCTAGAAAGCAGAAAGAGGTTGGAAGAAATGTTGGTGACAGACAGTT 5040
5041 GGAACAGAAAGCAGAGAGGGGAGGATCCAAAGATTCTGCAACATGTAGTGCATTTGGT 5100
5041 GGAACAGAAAGCAGAGAGGGGAGGATCCAAAGATTCTGCAACATGTAGTGCATTTGGT 5100
5101 TCTCTGGGTGACAAAGTTCCTCCAGGATAGGGCTGTAGAAAGGGGACAGGGGTGAGCC 5160
5101 TCTCTGGGTGACAAAGTTCCTCCAGGATAGGGCTGTAGAAAGGGGACAGGGGTGAGCC 5160
5161 AATGATTCAGTTAGGGGACACATCCAGCCAGGGTCTTGTCTGGCAAGCTAAAGAAATG 5220
5161 AATGATTCAGTTAGGGGACACATCCAGCCAGGGTCTTGTCTGGCAAGCTAAAGAAATG 5220
5221 AGAGCCCTCTAACCCTCCCTGAAAGTTAGGGGAGACAGAGACGTAAGAGATCCTTCTA 5280
5221 AGAGCCCTCTAACCCTCCCTGAAAGTTAGGGGAGACAGAGACGTAAGAGATCCTTCTA 5280
5281 GGGTAAGAGAGATGATCTGCTGACCAACATGGCTAGAGACAGAGATGGAGCCAG 5340
5281 GGGTAAGAGAGATGATCTGCTGACCAACATGGCTAGAGACAGAGATGGAGCCAG 5340
5341 TTACCCCTCAGAACCCAGCATCCCTCTTGCTTAAGAGAGGCTGGGCCCTTTCTGTTT 5400
5341 TTACCCCTCAGAACCCAGCATCCCTCTTGCTTAAGAGAGGCTGGGCCCTTTCTGTTT 5400
5401 AAGATCTTACTTTTCTCAGAGAGAGAGACAGACCTTTGTCCCTCCCTTTGGTCAA 5460
5401 AAGATCTTACTTTTCTCAGAGAGAGAGACAGACCTTTGTCCCTCCCTTTGGTCAA 5460
5401 AAGATCTTACTTTTCTCAGAGAGAGAGACAGACCTTTGTCCCTCCCTTTGGTCAA 5460
5461 TAAACACCCCTGTGTAAATTTAGTTATTTTACGTCAAGTTGCTCCAGAGACGTCCA 5520
5461 TAAACACCCCTGTGTAAATTTAGTTATTTTACGTCAAGTTGCTCCAGAGACGTCCA 5520
5461 TAAACACCCCTGTGTAAATTTAGTTATTTTACGTCAAGTTGCTCCAGAGACGTCCA 5520
5521 TCTGTAGACCTCTGCTCTTAACCTACCAAGGATATGGCCACATTCCTCACCAGAGAG 5580
5521 TCTGTAGACCTCTGCTCTTAACCTACCAAGGATATGGCCACATTCCTCACCAGAGAG 5580
5581 TGCNAAAGAGAGCCTTAGAGAAAGGTAACAGTAACAAAGATGGCCACAGAAATAAACA 5640
5581 TGCNAAAGAGAGCCTTAGAGAAAGGTAACAGTAACAAAGATGGCCACAGAAATAAACA 5640
5581 TGCNAAAGAGAGCCTTAGAGAAAGGTAACAGTAACAAAGATGGCCACAGAAATAAACA 5640
5641 ACTACTATCTTTTGAACCAAAATTTGTTTGTGAACAGAGAGGGGTGTGAGAGTAT 5700
5641 ACTACTATCTTTTGAACCAAAATTTGTTTGTGAACAGAGAGGGGTGTGAGAGTAT 5700
5641 ACTACTATCTTTTGAACCAAAATTTGTTTGTGAACAGAGAGGGGTGTGAGAGTAT 5700
5701 GTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 5760
5701 GTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 5760
5761 CTGGGGGACCTTTTATGCTAAAGAAATCTGATATTTGGCCCATATGCCAAAGGGGAT 5820
5761 CTGGGGGACCTTTTATGCTAAAGAAATCTGATATTTGGCCCATATGCCAAAGGGGAT 5820
5821 TGGGAGAGATCAGGCTTGCNAAACAGTAAGTGCNCAAGATGATTTGGGTGCTGAA 5880
5821 TGGGAGAGATCAGGCTTGCNAAACAGTAAGTGCNCAAGATGATTTGGGTGCTGAA 5880
5821 TGGGAGAGATCAGGCTTGCNAAACAGTAAGTGCNCAAGATGATTTGGGTGCTGAA 5880
5881 TCACCAAGGGGAGGCTGATCAGAGTGAAGAACATCAAGATTAAGCCACCTGTGGG 5940
5881 TCACCAAGGGGAGGCTGATCAGAGTGAAGAACATCAAGATTAAGCCACCTGTGGG 5940

Db 5881 TCACCAAGGGGAGGCTGATCAGAGTGAAGAACATCAAGATTAAGCCACCTGTGGG 5940
Qy 5941 GCTCAAGAGGAGCTTTTCAAGAGTGAAGGCCAAGCCATTTATTTATTCACAGATGAC 6000
Db 5941 GCTCAAGAGGAGCTTTTCAAGAGTGAAGGCCAAGCCATTTATTTATTCACAGATGAC 6000
Qy 6001 TCNAAATCAAACTGCAAGAGAGATTTAGTGTGAGAGATGGGCTGTCACTGTGGGACAC 6060
Db 6001 TCNAAATCAAACTGCAAGAGAGATTTAGTGTGAGAGATGGGCTGTCACTGTGGGACAC 6060
Qy 6061 TGACCTTGACATTTATTTAGTCACTAGGCCAAGAGACAGTACAGAGGGTACTGGGTCTA 6120
Db 6061 TGACCTTGACATTTATTTAGTCACTAGGCCAAGAGAGACAGTACAGAGGGTACTGGGTCTA 6120
Qy 6121 CTCACCTTGACAGGACAGTGTGAGAGATGGGTGACCTCATCTGATGAGAGGGGTGAG 6180
Db 6121 CTCACCTTGACAGGACAGTGTGAGAGATGGGTGACCTCATCTGATGAGAGGGGTGAG 6180
Qy 6181 CACCAACAGATCAAGTGTTCCTGTGTCTCATGCCAGAGATTCCTGGCAGTTTTCNAAAG 6240
Db 6181 CACCAACAGATCAAGTGTTCCTGTGTCTCATGCCAGAGATTCCTGGCAGTTTTCNAAAG 6240
Qy 6241 GACTAAGAGACTATCTCTGTGGAAACAAAGTATCCAGCCCTAAGCCCATTTTGTGCT 6300
Db 6241 GACTAAGAGACTATCTCTGTGGAAACAAAGTATCCAGCCCTAAGCCCATTTTGTGCT 6300
Qy 6301 AATTAATCAGAACCCCTGGGATGAGGCTCTGAGCAGACAGAGCTTTTAAAGATC 6360
Db 6301 AATTAATCAGAACCCCTGGGATGAGGCTCTGAGCAGACAGAGCTTTTAAAGATC 6360
Qy 6361 CCAGGTGATTCATGATCAGCAGCTGGAACAAACACAGCTCAGGTTCAACAGAAAGAGC 6420
Db 6361 CCAGGTGATTCATGATCAGCAGCTGGAACAAACACAGCTCAGGTTCAACAGAAAGAGC 6420
Qy 6420 AAGGTAGGAAAGCTTGGGATGGGAGGCTTCTTCCAGGACAGTATGAGAGGCTGTT 6480
Db 6420 AAGGTAGGAAAGCTTGGGATGGGAGGCTTCTTCCAGGACAGTATGAGAGGCTGTT 6480
Qy 6481 AGCAGTGTGGAGCTCTCTGCTGCTGATATAGCTATTCATCCATCCATCCATCCAT 6540
Db 6481 AGCAGTGTGGAGCTCTCTGCTGCTGATATAGCTATTCATCCATCCATCCATCCAT 6540
Qy 6541 ACACCCACCATCCATTTATGACACCATCTTCACATCCATCCATCCATCCATCCATCC 6600
Db 6541 ACACCCACCATCCATTTATGACACCATCTTCACATCCATCCATCCATCCATCCATCC 6600
Qy 6601 CCAGCATCCATCCAAACCTTCTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 6660
Db 6601 CCAGCATCCATCCAAACCTTCTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 6660
Qy 6661 CATTTATCCAAAGAGAACTGTATTTGCTAAATGTGGAGATTTAATTAATTTTAA 6720
Db 6661 CATTTATCCAAAGAGAACTGTATTTGCTAAATGTGGAGATTTAATTAATTTTAA 6720
Qy 6721 ACCTGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 6780
Db 6721 ACCTGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 6780
Qy 6781 TGCGAATCGGAGAAAGTTTGGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 6840
Db 6781 TGCGAATCGGAGAAAGTTTGGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 6840
Qy 6841 TTGAATCAAAATTAATCGGGCTGTGGCAAGTGTCTTTACACCGAGCCATTTTGTGATA 6900
Db 6841 TTGAATCAAAATTAATCGGGCTGTGGCAAGTGTCTTTACACCGAGCCATTTTGTGATA 6900
Qy 6901 CATCATTTATTTAAGAAAGCATCTTATATGATGATGATGATGATGATGATGATGATGAT 6960
Db 6901 CATCATTTATTTAAGAAAGCATCTTATATGATGATGATGATGATGATGATGATGATGAT 6960
Qy 6961 CCAGGATGACCTTTAACTCTGTCTTCCAGCTTCACCCGAGTGTAGTTTACAGGT 7020
Db 6961 CCAGGATGACCTTTAACTCTGTCTTCCAGCTTCACCCGAGTGTAGTTTACAGGT 7020
```

Db 6960 CCAAGATGACCTTTAACTCCTGCTTCCAGCCTCCACCCGAGCTAGGTTTACAGGT 7019  
 QY 7021 GTTCACTGGTGAATGCTTTTAAATCCAGCACTCTGTGGGGGGGGGGAGCGGATC 7080  
 Db 7020 GTTCACTGGTGAATGCTTTTAAATCCAGCACTCTGTGGGGGGGGGGAGCGGATC 7079  
 QY 7081 CCGAGTGGAGGCACTTTGGTCTACAGAGTTTCAGAGTATACCTGGGGGCTATACAGGAA 7140  
 Db 7080 CCGAGTGGAGGCACTTTGGTCTACAGAGTTTCAGAGTATACCTGGGGGCTATACAGGAA 7139  
 QY 7141 ACCCTATCCCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 7200  
 Db 7140 ACCCTATCCCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 7199  
 QY 7201 TTAGAGATATATAGTAGGATAGTAGGGCTGGTGAGGAGATCACTCTTTTGTATTT 7260  
 Db 7200 TTAGAGATATATAGTAGGATAGTAGGGCTGGTGAGGAGATCACTCTTTTGTATTT 7259  
 QY 7261 ATATAGTAAAGTACTACAAAGATGCAATATCTATCTATCTATCTATCTATCTATCTATC 7320  
 Db 7260 ATATAGTAAAGTACTACAAAGATGCAATATCTATCTATCTATCTATCTATCTATCTATC 7319  
 QY 7321 TATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATC 7380  
 Db 7320 TATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATC 7379  
 QY 7381 CTGCTTGAAGTCTGAATGCTCTCTATTTCTGGGTCAGCTCTTCACTCTAGTGTGGTTT 7440  
 Db 7380 CTGCTTGAAGTCTGAATGCTCTCTATTTCTGGGTCAGCTCTTCACTCTAGTGTGGTTT 7439  
 QY 7441 ACCAAGCCAGACATTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 7500  
 Db 7440 ACCAAGCCAGACATTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 7499  
 QY 7501 GGAGTCAGGGTCTGTGTCATGCTAAGCAAGCTCTGTGCAAGAGCTGCAAGCTGCAAGTCC 7560  
 Db 7500 GGAGTCAGGGTCTGTGTCATGCTAAGCAAGCTCTGTGCAAGAGCTGCAAGCTGCAAGTCC 7559  
 QY 7561 CCAATTTGTTCAGAGTGTGTCAGAGTGTGTCATATTTGCGAGGCTATATAGTCTCTC 7620  
 Db 7560 CCAATTTGTTCAGAGTGTGTCAGAGTGTGTCATATTTGCGAGGCTATATAGTCTCTC 7619  
 QY 7621 CACTCCCAAGTTCAGAGCACTTTGCGTCATCCAGTGGGGGGGCACTGTGTCTACCA 7680  
 Db 7620 CACTCCCAAGTTCAGAGCACTTTGCGTCATCCAGTGGGGGGGCACTGTGTCTACCA 7679  
 QY 7681 GTGCCCTGTCCCTGTCTTCCAGACCTAATATTTGCTGTGTAACAGTTCATGTAAATG 7740  
 Db 7680 GTGCCCTGTCCCTGTCTTCCAGACCTAATATTTGCTGTGTAACAGTTCATGTAAATG 7739  
 QY 7741 GGAATGGCTTCTGTATTTCTTTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 7800  
 Db 7740 GGAATGGCTTCTGTATTTCTTTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 7799  
 QY 7801 GGCAATGTGTCAGTGTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCT 7860  
 Db 7800 GGCAATGTGTCAGTGTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCT 7859  
 QY 7861 TGTGTGATTAACCACTTTCTGTTCATTTACTGATGGAATTTGTGGCCCAACCCCAAC 7920  
 Db 7860 TGTGTGATTAACCACTTTCTGTTCATTTACTGATGGAATTTGTGGCCCAACCCCAAC 7919  
 QY 7921 CCTTTTTTTTTTATTTAGACAAGGCTTTCTGTGTATCTTGAATCTTTGGCTGTCT 7980  
 Db 7920 CCTTTTTTTTTTATTTAGACAAGGCTTTCTGTGTATCTTGAATCTTTGGCTGTCT 7979  
 QY 7981 GACCTACCTGTAGACCAAGGCTGTGAGGCTGCTCCACTTTTGAACCTCGTGAC 8040  
 Db 7980 GACCTACCTGTAGACCAAGGCTGTGAGGCTGCTCCACTTTTGAACCTCGTGAC 8039  
 QY 8041 AGAGTAGCCATGAATCTCAAGACAAATTTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 8100  
 Db 8040 AGAGTAGCCATGAATCTCAAGACAA- TTTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 8098

QY 8101 TGTATGCGTGTATATGTGATGTTTGTGTCTTACAGGTGTCACATGTGTACTGTGTG 8160  
 Db 8099 TGTATGCGTGTATATGTGATGTTTGTGTCTTACAGGTGTCACATGTGTACTGTGTG 8158  
 QY 8161 TGGGACAGACAACCAACCGATGTGCCATTTCTCTGATGATACATGATCTTTTATATGTA 8220  
 Db 8159 TGGGACAGACAACCAACCGATGTGCCATTTCTCTGATGATACATGATCTTTTATATGTA 8218  
 QY 8221 TGTATATGATATGTTTATTTATTTAGTGTGCCAAGTATGAGGATATTTGTGTGAGTTTTCAC 8280  
 Db 8219 TGTATATGATATGTTTATTTAGTGTGCCAAGTATGAGGATATTTGTGTGAGTTTTCAC 8278  
 QY 8281 CTTCCTTTGTGGCTCTCCGATTTAACTAGCTCTCTGGGCTATGTAGCAATGCTTCA 8340  
 Db 8279 CTTCCTTTGTGGCTCTCCGATTTAACTAGCTCTCTGGGCTATGTAGCAATGCTTCA 8338  
 QY 8341 CTGATAGGCAATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 8400  
 Db 8339 CTGATAGGCAATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 8398  
 QY 8401 CGCACTGCACTGGCTTAAAGCTCACAAGTATCAGAGTGTAGCCAGGAGACACAG 8460  
 Db 8399 CGCACTGCACTGGCTTAAAGCTCACAAGTATCAGAGTGTAGCCAGGAGACACAG 8458  
 QY 8461 GGATATGCTGGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 8520  
 Db 8459 GGATATGCTGGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 8518  
 QY 8521 ATTTTAACTGATCTGAGATAGAGAGGCACTGTACCAATGGAGGCTCTTTTGT 8580  
 Db 8519 ATTTTAACTGATCTGAGATAGAGAGGCACTGTACCAATGGAGGCTCTTTTGT 8578  
 QY 8581 GTTTGGTTTGTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 8640  
 Db 8579 GTTTGGTTTGTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 8638  
 QY 8641 ATTAACATCTGTCTCAAAAAGCCATATAGAGTATGAGAGGCTTAAAGAGGCTT 8700  
 Db 8639 ATTAACATCTGTCTCAAAAAGCCATATAGAGTATGAGAGGCTTAAAGAGGCTT 8698  
 QY 8701 AAGCGGCTGTGATAGACACAGAGATAGCTGCACTATATAGCAAGCTTTGTTCAAA 8760  
 Db 8699 AAGCGGCTGTGATAGACACAGAGATAGCTGCACTATATAGCAAGCTTTGTTCAAA 8758  
 QY 8761 ACATGAGGAGGGGTATGTTTAAAGTGTGGCTGTGTAAACAGGCACTAAAGGAGCCAA 8820  
 Db 8759 ACATGAGGAGGGGTATGTTTAAAGTGTGGCTGTGTAAACAGGCACTAAAGGAGCCAA 8818  
 QY 8821 TGTAGACATTTGACTAAGAAAGATCAATCAAAAGCCGGGTGGGCAAGGATGAGGTTGG 8880  
 Db 8819 TGTAGACATTTGACTAAGAAAGATCAATCAAAAGCCGGGTGGGCAAGGATGAGGTTGG 8878  
 QY 8881 ACTACAGTGTCAAGACCCCATATAGAAAGCAAGTTCCTTCTCTGCTGCTGCTCAAG 8940  
 Db 8879 ACTACAGTGTCAAGACCCCATATAGAAAGCAAGTTCCTTCTCTGCTGCTGCTCAAG 8938  
 QY 8941 CTGGCTGCAAGGCGCATGCTCTCAATGCTTCTCTTATAGGCTGTCCACATG 8995  
 Db 8939 CTGGCTGCAAGGCGCATGCTCTCAATGCTTCTCTTATAGGCTGTCCACATG 8993

RESULT 3  
 AAF25499 standard; DNA: 23024 BP.  
 AAF25499:  
 AC AAF25499:  
 AC AAF25499:  
 DT 15-MAY-2001 (first entry)  
 XX Nucleotide sequence of a murine ABCA7 (ABCA7) transporter.  
 DE ABCA transporter; high-density lipoprotein cholesterol;  
 XX  
 KW

KW ABCA7 transporter; ABCA7 transporter; ss.  
 OS Mus musculus.  
 XX WO200109314-A1.  
 PN 08-FEB-2001.  
 XX 28-JAN-2000; 2000WO-FR00209.  
 XX 30-JUL-1999; 99FR-0009926.  
 XX (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.  
 PA (CNRS ) CNRS CENT NAT RECH SCI.  
 XX Chlunini G;  
 PI WPI; 2001-182953/18.  
 DR Selecting agents that modulate ABCA transporters, useful e.g. for  
 XX normalizing serum cholesterol levels, comprises using transgenic  
 PT animals with an inactive ABCA gene allele -  
 XX  
 PS Claim 12; Fig 22A; 113pp; French.  
 CC The specification describes a method for selecting or screening agents  
 CC that modulate ABCA transporters. The method comprises using non-human  
 CC recombinant mammals with an inactivated allele of the gene encoding the  
 CC ABCA transporter, or cells with an inactivated allele of ABCA  
 CC transporter, from any tissue of the recombinant mammal, preferably with  
 CC an allele truncated in one or both exons corresponding to the first  
 CC and/or second ATP-binding cassettes (NBD1 or 2). Compounds that stimulate  
 CC ABCA transporters may be useful for increasing (normalizing) serum  
 CC levels of high-density lipoprotein cholesterol. The present sequence  
 CC represents the genomic sequence of a murine ABCA7 (ABCA7) transporter.  
 CC  
 SQ Sequence 23024 BP; 5195 A; 6205 C; 6348 G; 5266 T; 10 other;  
 XX  
 XX  
 Query Match 1.7%; Score 153.2; DB 22; Length 23024;  
 Best Local Similarity 79.1%; Pred. No. 2e-26; Indels 0; Gaps 0;  
 Matches 182; Conservative 0; Mismatches 48;  
 QY 4552 ATAGAGCAGAGACTTGTCTTTAAGAAAAATGAAAGCCACAGTGTGGCAGACGCT 4611  
 DB 16839 ACACAGAGAAACCTGTCTAAAAAACCAAAACAGCGCGGACAGTGTGGCAGCT 16898  
 QY 4612 TTAATCCACGACTTGGGAGGAGCAGAGCAAGTTCGTGATTCAGAGCCAGCTGT 4671  
 DB 16899 TTAATCCACGACTTGGGAGGAGCAGAGCAAGTTCGTGATTCAGAGCCAGCTGT 16958  
 QY 4672 CTATAGAGTGTGATTCAGAGCAGAGGCTACAGAGAAACCTGTTTGAATAACCA 4731  
 DB 16959 CTACATGATGAGTTCAGAGCAGAGCAAGCTACAGAAAGAACCTGTTTGAATAACCA 17018  
 QY 4732 GAAAAACAAACAAACAAACAAACCAACCCAAACCCAAAC 4781  
 DB 17019 ACAAACAAACCAACCTTAAAAAACCAAAATAGTACAAACAAAC 17068  
 XX  
 XX  
 RESULT 4  
 AAL38337  
 ID AAL38337 standard; DNA; 215980 BP.  
 AC AAL38337;  
 XX  
 XX 15-AUG-2002 (first entry)  
 DE Complementary strand of a genomic sequence encoding a mouse Ngr3.  
 XX  
 XX Cerebroprotective; neuroprotective; cytosolic; Nogo receptor homologue;  
 KW Ngr3; Ngr3; axonal growth; central nervous system; CNS; cerebral injury;  
 KW spinal cord injury; stroke; demyelinating disease; multiple sclerosis;  
 KW monophasic demyelination; encephalomyelitis; Marchiafava-Bignami disease;

KW multifocal leukoencephalopathy; panencephalitis; Spongy degeneration;  
 KW Alexander's disease; Canavan's disease; metachromatic leukodystrophy;  
 KW Krabbe's disease; immune; bait protein; gene mapping; gene therapy;  
 KW transgenic animal; unregulated cellular growth; cancer; tumour; mouse;  
 KW murine; ds.  
 XX  
 XX Mus sp.  
 OS  
 XX WO200229059-A2.  
 PN 11-APR-2002.  
 PD 06-OCT-2001; 2001WO-US31488.  
 XX 06-OCT-2000; 2000US-238361P.  
 XX (UYTA ) UNIV YALE.  
 PA (BIO ) BIOGEN INC.  
 XX  
 XX Strittmatter SM, Cate RL, Sah DWJ;  
 PI WPI; 2002-416677/44.  
 DR Novel Nogo receptor homolog polypeptide, Ngr2 or Ngr3, useful for  
 XX treating central nervous system disorder, cerebral injury, spinal cord  
 PT injury, stroke, and demyelinating diseases -  
 XX  
 XX Example 16; Page 219-275; 277pp; English.  
 PS  
 CC The invention relates to a Nogo receptor homologue polypeptide, Ngr2 or  
 CC Ngr3, comprising a 50 amino acid LRKCT sequence, a 284 amino acid NLRKCT  
 CC sequence, or a 420, 461 or 392 amino acid sequence, all given in the  
 CC specification. The Ngr3 protein or its binding antibody is useful for  
 CC decreasing inhibition of axonal growth of a central nervous system (CNS)  
 CC neuron, by contacting the neuron Ngr3 or its antibody, and for treating  
 CC CNS disease, disorder or injury. Ngr3 or a vector comprising Ngr3 is  
 CC useful for treating cerebral injury, spinal cord injury, stroke,  
 CC demyelinating diseases, e.g. multiple sclerosis, monophasic  
 CC demyelination, encephalomyelitis, multifocal leukoencephalopathy,  
 CC panencephalitis, Marchiafava-Bignami disease, Spongy degeneration,  
 CC Alexander's disease, Canavan's disease, metachromatic leukodystrophy and  
 CC Krabbe's disease. Ngr3 is useful for inducing an immune response in a  
 CC mammal against Ngr3, as a bait protein in a two-hybrid or three-hybrid  
 CC assay, and as a research tool for identification, characterisation and  
 CC purification of interacting, regulatory proteins. The nucleotide  
 CC sequences of the invention are useful for screening for RFLP associated  
 CC with certain disorders, for genetic mapping, and for gene therapy. The  
 CC vector containing Ngr3 is useful for producing non-human transgenic  
 CC animals. The Ngr3 binding antibody is useful for isolating and purifying  
 CC Ngr3, for localisation and/or quantitation of Ngr3, and for diagnostic  
 CC and therapeutic purposes. The sequences of the invention, vectors and  
 CC antibodies are useful for treating or preventing unregulated cellular  
 CC growth such as cancer and tumour growth. This polynucleotide sequence  
 CC represents the  
 SQ Sequence 215980 BP; 56872 A; 50995 C; 50240 G; 54437 T; 3436 other;  
 XX  
 XX  
 Query Match 1.7%; Score 152.2; DB 24; Length 215980;  
 Best Local Similarity 82.2%; Pred. No. 1.1e-25; Indels 0; Gaps 0;  
 Matches 175; Conservative 0; Mismatches 38;  
 QY 4560 AGACTTGTCTTTAAGAAAAAATGAAGCCAGCAGTGTGGCAGGCTTTAATCCC 4619  
 DB 119710 ATATATTTTACTACATTAATAAAACAGCGGGGTGTGGCCATACCTTTATCC 119769  
 QY 4620 AGCACTTGGGAGGAGCAAGAGCAGAGATTTCTGAGTTCAAGCCAGCTGTCTATAG 4679  
 DB 119770 AGCACTCGGAGGAGCAGAGCGAGGATTTCTGAGTTGAGCCAGCTGTCTCAAG 119829  
 QY 4680 TGAGTTCAGAGCAGCAGGCGCTACAGAGAAACCTGTTTGAATAACCAAGCA 4739  
 DB 119830 TGAGTTCAGAGCAGCAGGCGCTACAGAGAAACCTGTCTGAAAAACCAAAAAA 119889

OY 4740 AACCAAAACAAACAAACCAACCA 4772  
DB 119890 AAAAAAAAAAAAAAAAACCACACA 119922

RESULT 5  
AAC82696  
ID AAC82696 standard; cDNA; 1445 BP.  
XX  
XX AAC82696;  
XX  
DT 15-MAR-2001 (first entry)  
XX  
XX Murine variant Zalpha32 cDNA.  
XX  
XX Zalpha32; murine; secreted alpha-helical protein-32; antiinflammatory;  
XX spermatogenesis; steroidogenesis; testicular differentiation;  
XX regulatory control; hypothalamic-pituitary-gonadal axis;  
XX reproductive system disorder; immunological system disorder; ss.  
XX  
XX Mus musculus.  
XX  
XX Key Location/Qualifiers  
XX FH 19.624  
XX CDS /\*tag= a  
XX FT /product= "variant Zalpha32"  
XX  
XX WO200071717-A1.  
XX  
XX 30-NOV-2000.  
XX  
XX 25-MAY-2000; 2000WO-US14563.  
XX  
XX 26-MAY-1999; 99US-0320159.  
XX  
XX (ZYMO ) ZYMOGENETICS INC.  
XX  
XX Conklin DC, Gao Z;  
XX  
XX WPI: 2001-032038/04.  
XX  
XX P-PSDB; AAB45760.  
XX  
XX Novel mammalian secreted alpha-helical protein-32 (Zalpha32). useful for  
XX identifying antagonists of its activity that are used in treating  
XX Zalpha32-induced inflammation  
XX  
XX Disclosure; Page 62-63; 73pp; English.

XX  
XX This invention describes a novel mammalian secreted alpha-helical  
XX protein-32 (Zalpha32) (I). The invention also describes (1) an isolated  
XX polynucleotide (II) encoding (1); (2) an antibody (III) that specifically  
XX binds to (1); and (3) an educational kit (IV) for the teaching of  
XX molecular biology and/or biochemistry comprising (II). The products of  
XX the invention have antiinflammatory activity. (III) as an antagonist to  
XX Zalpha32 is useful for treating Zalpha32-induced inflammation. Zalpha32  
XX can be used to identify inhibitors of its activity and to prepare  
XX antibodies that specifically binds to Zalpha32 epitopes. Antibodies or  
XX polypeptides, directly or indirectly conjugated to drugs or toxins are  
XX useful for in vivo diagnostic or therapeutic applications.  
XX Zalpha32-cytokine fusion proteins are useful for enhancing in vivo  
XX killing of target tissue. (I) and (II) are useful to identify and isolate  
XX receptors involved in spermatogenesis, steroidogenesis, testicular  
XX differentiation and regulatory control of the  
XX hypothalamic-pituitary-gonadal axis and also for testing disorders of the  
XX reproductive system and immunological systems.

XX  
XX Sequence 1445 BP: 418 A; 319 C; 398 G; 310 T; 0 other:  
XX  
XX Query Match 1.7%; Score 150.6; DB 22; Length 1445;  
XX Best Local Similarity 83.4%; Pred. No. 2.1e-26;  
XX Matches 171; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

OY 4570 TTTAAGCAAAAAATGAAGCCAGCGAGTGGCGACACGCTTAAATCCAGACCTTGGG 4629

DB 988 TTTAAAAAATGTGTGCTACCGGCGCTGTGGGCGACGCTTAAATCCAGACCTTGGG 1047  
OY 4630 AGGAGAGAGAGAGAGATTTCTGAGTTCAGAGCGCCGCTGATGAGTGTTCAG 4689  
DB 1048 AGGAGAGAGAGAGAGATTTCTGAGTTCAGAGCGCTGATGAGTGTTCAG 1107  
OY 4690 GACAGCCAGGCGTACACAGAGAAACCTGTTTGAACCAACCAACCAACCA 4749  
DB 1108 GACAGCCAGGCGTACAGAGAAACCTGTTTGAACCAACCAACCAACCA 1167  
OY 4750 AACCAAAACAAACCAACCAACCA 4774  
DB 1168 ACAACCAAAACCAACCAACCA 1192

RESULT 6  
AAA40866  
ID AAA40866 standard; DNA; 7208 BP.  
XX  
XX AAA40866;  
XX  
XX 16-AUG-2000 (first entry)  
XX  
XX  
XX Murine tumour necrosis factor alpha (TNFalpha) nucleotide sequence.  
XX  
XX Antisense oligonucleotide; phosphorothioate; TNFalpha; cytokine; inhibit;  
XX tumour necrosis factor alpha; inflammatory bowel disease; diabetes;  
XX rheumatoid arthritis; infectious disease; multiple sclerosis; hepatitis;  
XX pancreatitis; atopic dermatitis; allograft rejection;  
XX autoimmune disease; inflammatory disease; ds.  
XX  
XX Mus sp.  
XX  
XX WO200020645-A1.  
XX  
XX 13-APR-2000.  
XX  
XX 05-OCR-1999; 99WO-US23205.  
XX  
XX 05-OCR-1998; 98US-0166186.  
XX  
XX 18-MAY-1999; 99US-0313932.  
XX  
XX (ISIS-) ISIS PHARM INC.  
XX  
XX Baker BF, Bennett CF, Butler MW, Shanahan WJ;  
XX  
XX WPI: 2000-303808/26.  
XX  
XX P-PSDB; AAY88596.  
XX  
XX Oligonucleotide for treating diseases associated with human tumour  
XX necrosis factor-alpha (TNFalpha) such as, diabetes and rheumatoid  
XX arthritis, comprises nucleotide sequence complementary to intron of  
XX nucleic acid encoding TNFalpha -  
XX  
XX Example 8; Page 173-179; 283pp; English.

XX  
XX This sequence represents the murine tumour necrosis factor alpha  
XX (TNFalpha) nucleotide sequence. TNFalpha is an important cytokine that  
XX plays a role in host defence. It is produced mainly in macrophages and  
XX monocytes in response to infection, invasion, injury or inflammation.  
XX Overexpression of TNFalpha can result in disease states, particularly in  
XX infectious, inflammatory and autoimmune diseases. The invention relates  
XX to antisense oligonucleotides which are capable of modulating the  
XX TNFalpha gene expression. The oligonucleotides optionally have a  
XX phosphorothioate backbone, and may also optionally contain at least one  
XX 2'-O-methoxyethyl modification. The oligonucleotides are useful for  
XX modulating the expression of human TNFalpha in cells and tissues,  
XX reducing a human cell inflammatory response, reducing the blood glucose  
XX level in a human and treating a human having a disease or condition  
XX associated with TNFalpha. Examples of diseases associated with TNFalpha  
XX include diabetes, inflammatory bowel disease, multiple sclerosis,  
XX pancreatitis, rheumatoid arthritis, infectious disease, hepatitis, atopic

CC dermatitis or allograft rejection. The antisense oligonucleotides are  
CC also useful for modulating the function of a selected nucleic acid  
CC sequence in adipose tissue.

XX Sequence 7208 BP; 1750 A; 2011 C; 1810 G; 1637 T; 0 other;

Query Match 1.7%; Score 148.8; DB 21; Length 7208;  
Best Local Similarity 88.0%; Pred. No. 1.3e-25;  
Matches 162; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

QY 4574 AGAAAAATGAAACCCAGCGAGTGTGACACACGCTTTAATCCACAGCTTGGAGGC 4633  
DB 429 AAAAAAAAAAACTGGGAGTGTGACACACCTTTAATCCACAGCTTGGAGGC 488  
QY 4634 AGACAGGACGATTTCTGATTCAGAGTGTGACACGCTGTATAGAGTGTCCAGGACA 4693  
DB 489 AGAGCGAGCGGATTTCTGATTCAGAGTGTGACACGCTGTATAGAGTGTCCAGGACA 548  
QY 4694 GCCAGGCTACACAGAAACCCCTGTTTGAACCAACAAACAAACAAAC 4753  
DB 549 GCCAGGCTACACAGAAACCCCTGTTTGAACCAACAAACAAACAAAC 608  
QY 4754 AAAA 4757  
DB 609 AAAA 612

## RESULT 7

AAF63436/c  
ID AAF63436 standard; DNA; 6789 BP.

XX AAF63436;  
XX 14-MAY-2001 (first entry)

DE Murine CD39-L4 genomic DNA sequence.

XX Human CD39-like protein; apyrase; NDPase; platelet function inhibitor;  
KM myocardial infarction; cerebral ischaemia; angina; arterial thrombosis;  
KM cerebral artery thrombosis; platelet aggregation; inflammation;  
KM apoptosis; autoimmune disorder; neurological disorder;  
KM Alzheimer's disease; Parkinson's disease; cancer; CD39-L2; ds.

OS Mus sp.

PN WO200110205-A1.

XX 15-FEB-2001.

PF 09-AUG-2000; 2000WO-US21790.

XX 09-AUG-1999; 99US-0370265.  
PR 11-JAN-2000; 2000US-0481238.

PR 25-APR-2000; 2000US-0557800.

PR 26-MAY-2000; 2000US-0583231.

PR 30-JUN-2000; 2000US-0608285.

XX (HYSE-) HYSEQ INC.

PI Ford J, Mulero JJ, Yeung G;

DR WPI; 2001-147489/15.

XX Polynucleotides encoding human CD39-like polypeptides, with apyrase  
PT and/or NDPase activity, which are useful in the treatment of  
PT pathological conditions caused by thrombosis (e.g. myocardial  
PT infarction) and inflammatory disorders -  
XX  
XX Example 21; Page 108-110; 203pp; English.

CC This invention relates to polynucleotides encoding human CD39-like  
CC polypeptides with apyrase, and/or NDPase activity. The polypeptides having  
CC Apyrase, including NDPase, activity are useful for inhibiting platelet

CC function and can therefore be used in the prophylaxis or treatment of  
CC pathological conditions caused by or involving thrombosis or excessive  
CC coagulation or excessive platelet aggregation, such as myocardial  
CC infarction, cerebral ischaemia, angina, arterial thrombosis, cerebral  
CC artery thrombosis or intracardiac thrombosis, and conditions associated  
CC with venous thrombosis. CD39-L4 and CD39-L2 polypeptides are useful in  
CC modulating disease states (including platelet aggregation, inflammation  
CC and apoptosis) associated with ADP or other purinergic signalling by  
CC reducing the levels of NDPs. The polypeptides are also useful for  
CC prophylaxis or treatment of inflammation related disorders, such as  
CC disorders involving sepsis or systemic inflammatory response syndrome or  
CC SIRS (and associated conditions such as fever, tachycardia, tachypnea,  
CC cytokine overstimulation); autoimmune disorders such as thrombosis,  
CC atherosclerosis, acute pancreatitis, dermatitis, including psoriasis,  
CC cirrhosis, reperfusion injury, asthma, multiple sclerosis, arthritis;  
CC neurological disorders including neurodegenerative diseases, epilepsy,  
CC depression, Alzheimer's disease, Parkinson's disease, Huntington's  
CC disease, and amyotrophic lateral sclerosis; and cancer. The present  
CC sequence represents the murine CD39-L4 like protein CD39-L4 genomic DNA  
CC sequence.

XX Sequence 6789 BP; 1775 A; 1476 C; 1582 G; 1952 T; 4 other;

Query Match 1.6%; Score 145.6; DB 22; Length 6789;  
Best Local Similarity 62.3%; Pred. No. 8e-25;  
Matches 304; Conservative 0; Mismatches 164; Indels 20; Gaps 4;

QY 1854 GGAGGAGAGATGGCTCAGCTTCCAGACACTGCTGCTTGGACAGACCTAATTC 1913  
DB 1483 GGCCACAGAGATGACGAGTGAAGTATGCTCCCTGCTTACAGAGACTGATTC 1424  
QY 1914 AGTCCACAGGA--CTCATATGGTGGCTCAGAGCATGTGTAATCCAGAGGTT 1971  
DB 1423 AGTCCACAGCATGCTGAGAGGAGTGTGTAATGCTGCTTAATCCAGATCCAGAGATC 1364  
QY 1972 CCACACCTCTTCTGGCTCCACAGGACCCACATATGATACACAGCATATCATCAGG 2031  
DB 1363 TGACACCCCTCATGACAGTACTCACCCACACACCAATTAATAATTAATCTCAATTG 1304  
QY 2032 CAANAACCCATACACACATTAATAATTAAGAACTTAAAGGTCATGTTGGTAA 2091  
DB 1303 AAGAACTTAAGTATGATCATTCATTAATAATTAAGAACTTAAAGTCTGTTCTTT 1244  
QY 2092 CATTTGCTTACATGCTGATTTGAAGACATGTACACACACACTGAAGAGGATCTG 2151  
DB 1243 AAACATGTTTTTTTTTTTTTCTAGATAGACATTAATAAACAANAAG-----G 1190  
QY 2152 GGGCTGGAGAGATGGCTCAGCGGTTAAGAGACTGCTTCCGAAGGAGTCTTG 2211  
DB 1189 GGGCTGTCAGATGGCTCAGAGGGTAAAGACTGCTCTCTTTC-----GAAGGTCCTG 1134  
QY 2212 AGTTCAAATCTAGACACCAACCATGCTGCTCAACCATTCATTAAGATCTGACACCC 2271  
DB 1133 AGTTCAAATCTAGACACCAACCATGCTGCTCAACCATTCATTAAGATCTGACACCC 1074  
QY 2272 TCTTCTGTCATCTGACAGAGCTGACAGCTGACTACTTACTTACTTAATAAAT 2331  
DB 1073 TCTTTTGGTGTATCTAAGTC-----AGCTACAGTGTACTTATGTATAATAAAT 1022

QY 2332 AAATCTTT 2339

DB 1021 AAATCTTT 1014

AC AB226080;

DT 24-MAR-2003 (first entry)

DE Mouse DnaseX encoding genomic DNA SEQ ID NO 7.

XX Mouse; Dnase X; apoptosis; systemic lupus erythematosus; cystic fibrosis;  
 KW acquired; immune deficiency syndrome; AIDS; cancer; prostatic atrophy;  
 KW transgenic; dermatological; immunosuppressive; antiinflammatory;  
 KW anti-HIV; cytosolic; gene; ds.  
 XX  
 OS Mus musculus.  
 XX  
 PN EPI249495-A1.  
 XX  
 PD 16-OCT-2002.  
 XX  
 PE 11-APR-2001; 2001EP-0109024.  
 XX  
 PR 11-APR-2001; 2001EP-0109024.  
 XX  
 PA (DEKR-) DEUT KREBSFORSCHUNGSZENTRUM.  
 XX  
 PI Not given;  
 XX  
 DR WPI: 2003-042057/04.  
 XX  
 PT New DNA encoding murine DnaseX and related vectors, proteins and  
 PT antibodies useful for treating diseases that involve apoptosis, e.g.,  
 PT systemic lupus erythematosus,  
 PS  
 PS Claim 2; Fig 1; 72pp; German.  
 XX  
 XX The invention relates to a DNA (I) that: (i) (AB226077-AB226079) encodes  
 CC any of three splice forms of murine DnaseX (II) of 295, 295 and 61 amino  
 CC acids (ABP60403-ABP60405); (ii) is a 40116 bp genomic sequence  
 CC (AB226080); or (iii) encodes a protein (IIta) with the biological  
 CC properties of (II) and is a fragment or allelic or other variant of (I)  
 CC or (Ii). Dnase X degrades DNA inducing apoptosis, immunogenic DNA  
 CC complexes inhibiting the immune response in systemic lupus erythematosus  
 CC and mucus in the lungs of cystic fibrosis patients. (Ii) vectors  
 CC containing (I) and the encoded proteins (II), are used to diagnose,  
 CC prevent or treat diseases associated with apoptosis, specifically  
 CC systemic lupus erythematosus, acquired immune deficiency syndrome,  
 CC cancer, cystic fibrosis and prostatic atrophy. Also (I) and (II) can be  
 CC used to screen for specific binding partners, potential therapeutic  
 CC agents. Non-human animals in which (I) is altered (particularly  
 CC inactivated) are used to study the specified diseases and/or to  
 CC characterise genes and/or for testing potential therapeutic agents. The  
 CC present sequence is that of mouse DnaseX encoding genomic DNA sequence of  
 CC  
 XX  
 SQ Sequence 40116 BP; 11523 A; 9280 C; 9062 G; 10251 T; 0 other;  
 Query Match 1.68; Score 145.2; DB 25; Length 40116;  
 Best Local Similarity 81.68; Pred. No. 2.5e-24;  
 Matches 168; Conservative 0; Mismatches 38; Indels 0; Gaps 0;  
 QY 4569 CTTTAAAGAAAATGAAGCCAGCAGTGTGGACACGCTTTATCCAGCACTTGG 4628  
 DB 32014 CGTGACATATAAAAAACAGCGCGCGGTGTGGCATGCTTTATCCAGCACTCGG 32073  
 QY 4629 GAGGCAAGACGACGAGATTCTGAGTTCAAGGCCAGCTGTATAGAGTAGTTCCA 4688  
 DB 32074 GAGGCAAGACGACGAGATTCTGAGTTGAGGCCAGCTGTCTACAGAGTAGTTCCA 32133  
 QY 4689 GAGCAGCGAGGCTACAGAGAAACCCCTTTTGAAGAAAAACGAAAAACAAAC 4748  
 DB 32134 GAGCAGCGAGGCTATATAAGAAACCCCTGTCTCAAAAAACCAAAAAA 32193  
 QY 4749 AAAACAAAACAAACCAACCAAA 4774  
 DB 32194 CAAAACAAAACCAACCAAAACCA 32219  
 RESULT 9  
 ABS56565/c  
 ID ABS56565 standard; DNA: 171936 BP.

XX  
 AC ABS56565;  
 XX  
 DT 20-JAN-2003 (first entry)  
 XX  
 DE Human SUF2 genomic DNA sequence.  
 XX  
 KW SUF2; ds; glucosamine-6-sulphatase; cancer; ischemia; human;  
 KW tumour; angiogenesis; coronary; carotid; arterial occlusive disease;  
 KW peripheral arterial disease; atherosclerosis; myointimal hyperplasia;  
 KW thromboangitis obliterans; thrombotic disorder; vasculitis;  
 KW heart attack; myocardial infarction; vascular death; inflammation;  
 KW rheumatoid arthritis; asthma; adult respiratory distress syndrome;  
 KW sarcoidosis; hypersensitivity pneumonitis; multiple sclerosis;  
 KW allograft rejection; lymphoma; thrombosis; sulphatase;  
 KW chromosome 20q12-13.2.  
 XX  
 OS Homo sapiens.  
 XX  
 PN W0200259327-A2.  
 XX  
 PD 01-AUG-2002.  
 XX  
 PE 26-DEC-2001; 2001WO-US49793.  
 XX  
 PR 27-DEC-2000; 2000US-258577P.  
 XX  
 PR 09-FEB-2001; 2001US-267831P.  
 XX  
 PA (REGC ) UNIV CALIFORNIA.  
 XX  
 PI Rosen SD, Hemmerich S, Tomita M, Palmeri D;  
 XX  
 DR WPI: 2002-636541/68.  
 XX  
 PT New sulfatase polypeptides having glucosamine-6-sulphatase activity,  
 PT useful in screening, discovery and preparation of diagnostic and  
 PT therapeutic agents for treating cancer, ischemic conditions,  
 PT inflammation, or thrombosis  
 XX  
 XX Disclosure; Page 248-293; 293pp; English.  
 XX  
 XX This invention relates to the DNA and protein sequences of a novel  
 CC polypeptide having glucosamine-6-sulphatase activity. The sulphatases of  
 CC are useful in screening, discovery and preparation of diagnostic and  
 CC therapeutic agents for treating cancer, ischaemic conditions,  
 CC inflammation, or thrombosis. The nucleic acids are useful in preparing  
 CC the sulfatase polypeptides, identifying the expression of genes in a  
 CC biological specimen, or generating transgenic non-human animals or  
 CC site-specific gene modification in cell lines. The host cells are  
 CC useful in replicating and/or expressing the polynucleotides or nucleic  
 CC acids. The agents are useful in treating the disorders cited above by  
 CC reducing tumour growth, inflammation, and thrombosis, or increasing  
 CC angiogenesis, e.g. by treating coronary, carotid, or arterial occlusive  
 CC disease, peripheral arterial disease, atherosclerosis, myointimal  
 CC hyperplasia, thromboangitis obliterans, thrombotic disorders,  
 CC vasculitis; or preventing ischemic conditions, heart attack (myocardial  
 CC infarction), or other vascular death. The sulphatases and/or agents are  
 CC also useful in treating rheumatoid arthritis, asthma, adult respiratory  
 CC distress syndrome, sarcoidosis, hypersensitivity pneumonitis, multiple  
 CC sclerosis, allograft rejection, and spread of lymphomas to cutaneous  
 CC sites. The present sequence represents the human SUF2 genomic DNA  
 CC sequence of the invention. This gene maps to human chromosome  
 CC 20q12-13.2.  
 XX  
 SQ Sequence 171936 BP; 47285 A; 42171 C; 41180 G; 41300 T; 0 other;  
 Query Match 1.68; Score 144.8; DB 24; Length 171936;  
 Best Local Similarity 83.78; Pred. No. 6.0e-24;  
 Matches 164; Conservative 0; Mismatches 32; Indels 0; Gaps 0;  
 QY 4567 GTCTTAAGAAAATGAAGCCAGCAGTGTGGACACGCTTTATCCAGCACTT 4626  
 DB 17076 GCCTTTAAAAAACTATCCGCGCGGTGTGGCGCAAGCTTTAGTCCAGCACTT 17017









KW EHD2; endocytosis; IGFI signaling; suppressing adhesion; invasion;  
 KW metastasis; bone formation; osteoporosis; mouse; murine; gene; ds.  
 XX  
 OS Mus sp.  
 XX  
 PN US2002115069-A1.  
 XX  
 PD 22-AUG-2002.  
 XX  
 PF 17-MAY-1999; 99US-0312762.  
 XX  
 PR 21-FEB-1997; 97IL-0120283.  
 PR 20-FEB-1998; 98US-0026898.  
 XX  
 PA (HORO/) HOROWITZ M.  
 PA (MINT/) MINTZ L.  
 XX  
 PI Horowitz M, Mintz L;  
 DR WPI; 2002-731288/79.  
 XX  
 PT Novel recombinant Eps15 homology domain containing protein useful for  
 PT regulating endogenous EHD protein by regulating insulin-like growth  
 PT factor I receptor cell signaling via altered clathrin coated pit  
 PT mediated endocytosis  
 XX  
 PS Claim 6: Page 32-39; 88pp; English.  
 XX  
 CC The invention relates to a novel recombinant Eps15 homology (EH) domain  
 CC containing protein 1 or 2 (EHD1 or EHD2) with an N-terminal region  
 CC containing a nucleotide binding consensus site, a central coiled coil  
 CC structure and a C-terminal region including an EH domain, where the  
 CC polypeptide participates in endocytosis. The peptides, peptide analogues  
 CC and/or the polynucleotide sequences are useful for regulating EHD1/2  
 CC protein activity in vivo, where the EHD1/2 protein and its gene are  
 CC useful for upregulating the protein activity. The novel method is useful  
 CC for lowering the rate of IGFI signaling and therefore useful for  
 CC suppressing adhesion, invasion and metastasis where EHD is overexpressed,  
 CC and therefore endocytosis. The method is also useful for elongating IGFI  
 CC effects and therefore for increasing bone formation in osteoporosis  
 CC conditions. This polynucleotide sequence represents the genomic DNA of a  
 CC mouse EHD1 gene of the invention.  
 XX  
 SQ Sequence 14707 BP; 3397 A; 3826 C; 3894 G; 3568 T; 22 other;  
 XX  
 QY Query Match 1.6%; Score 142.6; DB 24; Length 14707;  
 Best Local Similarity 75.1%; Pred. No. 6.5e-24;  
 Matches 178; Conservative 0; Mismatches 59; Indels 0; Gaps 0;  
 OY 4576 AAAAAAATGAAGCCAGCAGTGTGGCAGACGCTTTAATCCAGCACTTGGAGGCAG 4635  
 DB 12047 AAGAAATTAAGCGGGGGGGGATGGCAGACGCTTTAATCCAGCACTTGGAGGCAG 12106  
 OY 4636 AAGCAGCAGATTTCTGAGTTAAAGCCAGCCTGCTATAGAGTGAAGTCCAGACAGC 4695  
 DB 12107 AAGCGGGCGGATTTCTGAGTTGAAGCCAGCCTGCTATAGAGTGAAGTCCAGACAGC 12166  
 OY 4696 CAGGGCTACACAGAGAAACCTGTTGAAAAACGAAACCAAAACCAAAACCA 4755  
 DB 12167 CAGGGCTACACAGAGAAACCTGTTGAAAAACGAAACCAAAACCAAAACCA 12226  
 OY 4756 AACCAAAACCAAAACCAAAACCTGCTATAGAGTGAAGTCCAGACAGC 4812  
 DB 12227 GTTTGAGACCAAGTGAAGCCCTCCCTTGTGCTGAGCCTGAGTGTGGGTTT 12283

XX Mouse high growth region.  
 DE  
 XX  
 KW High growth region; high growth phenotype; Socs2; body size;  
 KW suppressor of cytokine signaling 2; ds; mouse.  
 XX  
 OS Mus sp.  
 XX  
 PN US2002155564-A1.  
 XX  
 PD 24-OCT-2002.  
 XX  
 PF 26-JAN-2001; 2001US-0771208.  
 PF 29-DEC-1997; 97US-0999477.  
 PR  
 XX (REGC ) UNITV CALIFORNIA.  
 PA  
 PI Medrano JF, Bradford E, Horvat S;  
 DR WPI; 2003-182637/18.  
 XX  
 PT Novel gene that when downregulated or knocked-out, results in high  
 PT growth phenotype, useful for regulating body size in mammals e.g.  
 PT rodent, bovine and canine  
 XX  
 PS Disclosure; SEQ ID NO 18; 49pp; English.  
 XX  
 CC The invention describes an isolated nucleic acid molecule encoding a  
 CC gene product that, when knocked out, results in a high growth (hg)  
 CC phenotype. For example a nucleic acid disrupting the Socs2 gene is  
 CC useful for producing an animal characterised by a hg phenotype, by  
 CC inhibiting expression of Socs2 (suppressor of cytokine signaling 2)  
 CC gene. The nucleic acids of the invention are useful for regulating  
 CC body size in mammals. The nucleic acids of the invention  
 CC are useful for regulating body size in mammals. This sequence  
 CC represents the mouse high growth region.  
 CC Note: This sequence did not form part of the printed specification  
 CC but was obtained in electronic format directly from the US patent  
 CC office at  
 CC seqdata.uspto.gov/sequence.html?docID=20020155564.  
 XX  
 SQ Sequence 659158 BP; 177521 A; 147222 C; 149414 G; 184754 T; 247 other;  
 XX  
 QY Query Match 1.6%; Score 142.2; DB 25; Length 659158;  
 Best Local Similarity 85.0%; Pred. No. 5.8e-23;  
 Matches 159; Conservative 0; Mismatches 28; Indels 0; Gaps 0;  
 OY 4576 AAAAAAATGAAGCCAGCAGTGTGGCAGACGCTTTAATCCAGCACTTGGAGGCAG 4635  
 DB 20987 AAGAAAGAAAGCGGGGGGGGATGGCAGACGCTTTAATCCAGCACTTGGAGGCAG 21046  
 OY 4636 AAGCAGCAGATTTCTGAGTTAAAGCCAGCCTGCTATAGAGTGAAGTCCAGACAGC 4695  
 DB 21047 AAGCGGGCGGATTTCTGAGTTGAAGCCAGCCTGCTATAGAGTGAAGTCCAGACAGC 21106  
 OY 4696 CAGGGCTACACAGAGAAACCTGTTGAAAAACGAAACCAAAACCAAAACCA 4755  
 DB 21107 CAGGGCTATACAGAGGAAACCTGCTGCAAAAGCAACCAAAACCAAAACCA 21166  
 OY 4756 AACCAAA 4762  
 DB 21167 AAAAAAA 21173

Mouse Ret15 genomic DNA sequence.

KM Ret ligand 5; RetL5; autophosphorylation; tumour; renal; nephrotropic;  
KM Alzheimer's disease; Parkinson's disease; Huntington's disease; mouse;  
KM valine; nootropic; anti-HIV; neuroprotective; antibacterial; ds;  
KM cerebroprotective; hemostatic; antiinflammatory; antiviral; neuroleptic.

OS Mus sp

PN WO200116169-A2.

PD 08-MAR-2001.

PF 01-SEP-2000; 2000WO-US24111.

PR 01-SEP-1999; 99US-0152024.

PA (BIOJ ) BIOGEN INC.

PI Worley D.

DR WPI; 2001-235091/24.

PT Novel Ret ligand polypeptide useful for suppressing growth of a tumor  
PT cell that expresses Ret and for modulating Ret signal transduction  
PT involving a cell expressing Ret polypeptide or Ret ligand polypeptide

PS Claim 1; Page 42-53; 76pp; English

CC The Invention relates to mouse and human Ret Ligand 5 (RetL5)

Sequence 15914 BP; 3898 A; 3883 C; 4064 G; 4064 T; 5 other;

Query Match	1.68;	Score 141.8;	DB 22;	Length 15914;
-------------	-------	--------------	--------	---------------

Matches 161; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

4562 ACTTTGCTTTAAGAAATAAGCCAGCAGTGGTGGCACACGCCCTTTAATCCAG 4621

Db 4668 ATTTCTTTTAAAAAGAGCCGGCGATGGTGGCACATGCCCTTAATCCAG 4727

4622 CACTTGGGAGCAGAAGCAGGATTTCTGAGTTC AAGGCCAGCCTGGTCTATAGAGTG 4681

Db 4728 CACTTGGGAGGCAGAGGCAGATTCTGAGTTCAGGCCAGCCTGGACTACAGAGTG 4787

4682 AGTTCAGGACAGCCAGGGCTACACAGAGAAACCTGTTTGAAAAACCGAAAAACAA 4741

Db 4788 AGTTCAGGACAGCCAGGCTACTCAGAGAACTCTGTCTTGGAATAAAAGAGGAA 4847

7

QY 4742 ACACAAACAAACA 4754

Db 4848 GAAAGGAAAAAGA 4860

Search completed: September 13, 2003, 19:46:15  
Job time : 2045 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

## OM nucleic - nucleic search, using sw model

Run on: September 13, 2003, 19:05:48 ; Search time 31328 Seconds  
(without alignments)  
11746.109 Million cell updates/sec

Title: US-09-877-935-1

Perfect score: 8995

Sequence: 1 gatctgtgtcaccaagaca.....ctctagctctgcaccatg 8995

## Scoring table:

IDENTITY\_NUC  
Gapop 10.0 , Gapept 1.0

Searched: 2888711 segs, 20454813366 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

## Database :

GenBml:.\*  
1: gb\_ba:.\*  
2: gb\_hlg:.\*  
3: gb\_in:.\*  
4: gb\_om:.\*  
5: gb\_ov:.\*  
6: gb\_pat:.\*  
7: gb\_ph:.\*  
8: gb\_pl:.\*  
9: gb\_pr:.\*  
10: gb\_ro:.\*  
11: gb\_sts:.\*  
12: gb\_sy:.\*  
13: gb\_un:.\*  
14: gb\_vl:.\*  
15: em\_ba:.\*  
16: em\_fun:.\*  
17: em\_hum:.\*  
18: em\_in:.\*  
19: em\_mu:.\*  
20: em\_om:.\*  
21: em\_or:.\*  
22: em\_ov:.\*  
23: em\_pat:.\*  
24: em\_ph:.\*  
25: em\_pl:.\*  
26: em\_ro:.\*  
27: em\_sts:.\*  
28: em\_un:.\*  
29: em\_vl:.\*  
30: em\_hlg\_hum:.\*  
31: em\_hlg\_inv:.\*  
32: em\_hlg\_other:.\*  
33: em\_hlg\_mus:.\*  
34: em\_hlg\_pln:.\*  
35: em\_hlg\_rod:.\*  
36: em\_hlg\_man:.\*  
37: em\_hlg\_vrt:.\*  
38: em\_sy:.\*  
39: em\_higo\_hum:.\*  
40: em\_higo\_mus:.\*  
41: em\_higo\_other:.\*

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	8995	100.0	8995	6	AX027824
2	8762.2	97.4	192060	2	AC098570
3	6067	67.4	170074	2	AC132885
4	3116.8	34.7	193472	2	AC115173
5	3016.2	33.5	149109	2	AC101018
6	564.6	6.3	779	11	BV067438
7	188.6	2.1	203605	2	AC119514
8	188.6	2.1	234549	2	AC128412
9	188.6	2.1	240435	2	AC094169
10	188.4	2.1	262322	2	AC132539
11	188.2	2.1	238330	2	AC114361
12	188	2.1	149425	2	AC013320
13	188	2.1	191754	9	AC021016
14	186.6	2.1	568	11	G76462
15	185	2.1	195636	2	AC100085
16	184.8	2.1	1220138	2	AC115965
17	184	2.0	184591	2	AC102777
18	183.6	2.0	273800	10	AF100956
19	181.8	2.0	193534	2	AC130278
20	179.2	2.0	93419	10	AL929418
21	179.2	2.0	218558	2	AC025964
22	179.2	2.0	240536	10	AC025910
23	179.2	2.0	241400	10	AL646093
24	178.2	2.0	198923	2	AC109263
25	177.8	2.0	209206	2	AC117095
26	177.8	2.0	246991	2	AC122956
27	177.6	2.0	268276	2	AC132103
28	177.4	2.0	192129	2	AC134667
29	177.4	2.0	208161	2	AC074145
30	177.2	2.0	166167	10	AL611936
31	177.2	2.0	248677	2	AL844206
32	177	2.0	210372	10	AC122807
33	177	2.0	224766	2	AC132138
34	176.6	2.0	173425	2	AC131034
35	176.6	2.0	204316	10	AC087541
36	176.2	2.0	180526	2	AC136083
37	175.8	2.0	200057	2	AC073810
38	175.6	2.0	227423	2	AC134604
39	175.6	2.0	251450	2	AC133083
40	174.4	1.9	207814	10	AL590994
41	174.2	1.9	141166	10	AC122423
42	174	1.9	170142	2	AC140286
43	174	1.9	173567	2	AC120412
44	174	1.9	187376	2	AC138613
45	174	1.9	231350	2	AC123357

## ALIGNMENTS

RESULT 1  
AX027824 LOCUS 8995 bp DNA linear PAT 16-SEP-2000  
DEFINITION Sequence 1 from Patent WO0034492.  
ACCESSION AX027824  
VERSION AX027824.1 GI:10188668  
KEYWORDS  
SOURCE  
ORGANISM Mus sp.  
Mus sp.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 Robine,S., Louvard,D., Pinto,D. and Jaisser,F.  
AUTHORS Regulatory sequences of the mouse villin gene - use in transgenesis  
TITLE Patent: WO 0034492-A 1 15-JUN-2000;  
JOURNAL

Pred. No. is the number of results predicted by chance to have a





QY	4141	TCACCTGAGTTGGGAGAAATCTGCTTGGGGGCTTCCAGGTCYTTGTTAGCAGAGGGT	4200
Db	4141	TCACCTGAGTTGGGAGAAATCTGCTTGGGGGCTTCCAGGTCYTTAGCAGAGGGT	4200
QY	4201	ATCCTTGTATAGGCGATGACCTAGTCTATGTTGTACTACATTCCTCCAGTTAAAG	4260
Db	4201	ATCCTTGTATAGGCGATGACCTAGTCTATGTTGTACTACATTCCTCCAGTTAAAG	4260
QY	4261	CTGGAATAAAACCCAGCGAGCCCGCAGATTCTCTACAGTTGTACCCCAAGAACA	4320
Db	4261	CTGGAATAAAACCCAGCGAGCCCGCAGATTCTCTACAGTTGTACCCCAAGAACA	4320
QY	4321	AGACAGTATGATATGCAAGGATAGTACGTGGGAGAGAAGAACTTAAACCCCAAG	4380
Db	4321	AGACAGTATGATATGCAAGGATAGTACGTGGGAGAGAAGAACTTAAACCCCAAG	4380
QY	4381	GCCCAAGGTTCCGTTCCCTAGTTCCACAATGCCAGTATAGAGTGCCTACTATGGCTG	4440
Db	4381	GCCCAAGGTTCCGTTCCCTAGTTCCACAATGCCAGTATAGAGTGCCTACTATGGCTG	4440
QY	4441	TGAGTTGGTACCTACAAACATGATGATGTTCAATGTGTAGTGTATTAATCTGACAC	4500
Db	4441	TGAGTTGGTACCTACAAACATGATGATGTTCAATGTGTAGTGTATTAATCTGACAC	4500
QY	4501	TTGGGAGGCTAGACGAGGAGATTGCTATATGTTTGAGGCCAGCCTGAGCTATAGACGA	4560
Db	4501	TTGGGAGGCTAGACGAGGAGATTGCTATATGTTTGAGGCCAGCCTGAGCTATAGACGA	4560
QY	4561	GACTTTGTCTTAAAGAAAAATGAAGCAAGCAGTGTGGGCAACGCCCTTAAATCCCA	4620
Db	4561	GACTTTGTCTTAAAGAAAAATGAAGCAAGCAGTGTGGGCAACGCCCTTAAATCCCA	4620
QY	4621	GCACCTTGGGAGGAGAAACAGGAGATTCTGAGTTCAAGGCCAGCCTGTCTATAGAT	4680
Db	4621	GCACCTTGGGAGGAGAAACAGGAGATTCTGAGTTCAAGGCCAGCCTGTCTATAGAT	4680
QY	4681	GAGTTCCAGGACAGCCAGGGCTACACAGAGAAACCCCTTTTGAAGAAACAGAAAAACA	4740
Db	4681	GAGTTCCAGGACAGCCAGGGCTACACAGAGAAACCCCTTTTGAAGAAACAGAAAAACA	4740
QY	4741	AACAAAAAATAAAACAAAAACCAACCCAAACCCCAACCTCTCATCTCTCATCTCTC	4800
Db	4741	AACAAAAAATAAAACAAAAACCAACCCAAACCCCAACCTCTCATCTCTCATCTCTC	4800
QY	4801	TAGCGTGTCTGTCTAGTGTGTAGAGTTTGGGGAATTCAAGCTTATATTAATAATAGCC	4860
Db	4801	TAGCGTGTCTGTCTAGTGTGTAGAGTTTGGGGAATTCAAGCTTATATTAATAATAGCC	4860
QY	4861	TTTTTATCTGTCTAGAGACGAGAAAGTTTCACTCTGGGACACAGTGGGCCCTTGAGA	4920
Db	4861	TTTTTATCTGTCTAGAGACGAGAAAGTTTCACTCTGGGACACAGTGGGCCCTTGAGA	4920
QY	4921	AAGTACTCTTGGCAGGCCCAAAATTTGGGAAGGCTTCTGGAGGAAGTGTGCCGAT	4980
Db	4921	AAGTACTCTTGGCAGGCCCAAAATTTGGGAAGGCTTCTGGAGGAAGTGTGCCGAT	4980
QY	4981	CAGACTACTGTTCTAGAAAGCAGAAAGAGGTTTGAAGAAATGTTGTGTGACAGACACTT	5040
Db	4981	CAGACTACTGTTCTAGAAAGCAGAAAGAGGTTTGAAGAAATGTTGTGTGACAGACACTT	5040
QY	5041	GGAACAGAGAGACAGGAGGGGAGGCATCAAGATTCGAACATTTAGCTGACTTTTGGT	5100
Db	5041	GGAACAGAGAGACAGGAGGGGAGGCATCAAGATTCGAACATTTAGCTGACTTTTGGT	5100
QY	5101	TCTCGGGGTGAAGTGTCCCAAGGGATAGGGGCTGTGAAGAAAGGAGACAGGGGGTACCC	5160
Db	5101	TCTCGGGGTGAAGTGTCCCAAGGGATAGGGGCTGTGAAGAAAGGAGACAGGGGGTACCC	5160
QY	5161	AATGAGTTCAAGTTGAGGGACACATCCAGCCAGGGTCTTGTGTGGCAAGCTAAAGAAATG	5220
Db	5161	AATGAGTTCAAGTTGAGGGACACATCCAGCCAGGGTCTTGTGTGGCAAGCTAAAGAAATG	5220

---

QY	5221	AGAGCCCTCTAACCTCCCTGTAAGTTTAGGGAGACAGAGAGCTGAGAGATCCTCTTA	5280
Db	5221	AGAGCCCTCTAACCTCCCTGTAAGTTTAGGGAGACAGAGAGCTGAGAGATCCTCTTA	5280
QY	5281	GGGTGAAGAGAGGATGTGCTGTGACCAATGCTAGGAGCGAAGCAAGTGTGACAG	5340
Db	5281	GGGTGAAGAGAGGATGTGCTGTGACCAATGCTAGGAGCGAAGCAAGTGTGACAG	5340
QY	5341	TTACCCCTCAGAACAGCCATCCCTCTTGGCTCTAAGAGAGCTGGGCCCTTTCTGTTT	5400
Db	5341	TTACCCCTCAGAACAGCCATCCCTCTTGGCTCTAAGAGAGCTGGGCCCTTTCTGTTT	5400
QY	5401	AAGAATCTTACTTTTCTTACAGAGAGCGACAGCCTTTGTCCCTCTGTTGTCAA	5460
Db	5401	AAGAATCTTACTTTTCTTACAGAGAGCGACAGCCTTTGTCCCTCTGTTGTCAA	5460
QY	5461	TAAACAGCCGTGTGTAAACATTAATTAATTTACTGTACAGTTTCTCAGAGACGTCCA	5520
Db	5461	TAAACAGCCGTGTGTAAACATTAATTAATTTACTGTACAGTTTCTCAGAGACGTCCA	5520
QY	5521	TCTGTAGACCTGTGCTCTTAACCTCACCAAGGTATGGCCACATTTCTCACCCAGAAAG	5580
Db	5521	TCTGTAGACCTGTGCTCTTAACCTCACCAAGGTATGGCCACATTTCTCACCCAGAAAG	5580
QY	5581	TGCAGAGAGAGCCTTAGAGAAAGGTTAACAGTAAACAAAGATGCCACAAATTAACAAA	5640
Db	5581	TGCAGAGAGAGCCTTAGAGAAAGGTTAACAGTAAACAAAGATGCCACAAATTAACAAA	5640
QY	5641	ACTACTATCTTGTATACCAATTTGTTTGTGTGTAACCAAGAGGGGGTGTGAGTGTAT	5700
Db	5641	ACTACTATCTTGTATACCAATTTGTTTGTGTGTAACCAAGAGGGGGTGTGAGTGTAT	5700
QY	5701	GT	5760
Db	5701	GT	5760
QY	5761	CTTGGGGACATTTTATCTGTAAGAAATATCTGATATTTGGCCGCCATGCCAACAGGGGAT	5820
Db	5761	CTTGGGGACATTTTATCTGTAAGAAATATCTGATATTTGGCCGCCATGCCAACAGGGGAT	5820
QY	5821	TGGGAGAGTACAGGCTTCTGCAACACAGTAAGTGGCCAAAGATGGATGGGGCTTAA	5880
Db	5821	TGGGAGAGTACAGGCTTCTGCAACACAGTAAGTGGCCAAAGATGGATGGGGCTTAA	5880
QY	5881	TCACCAAGGGGAGGCTGATCAGAGTGACAGAACATCAACAAGATTAAGCCACCTGTGGG	5940
Db	5881	TCACCAAGGGGAGGCTGATCAGAGTGACAGAACATCAACAAGATTAAGCCACCTGTGGG	5940
QY	5941	GCTCAGAGAGGAGTTTACAAGAGGTAAAGGCCAAGCCATTTATTAATCCAAAGACATGAC	6000
Db	5941	GCTCAGAGAGGAGTTTACAAGAGGTAAAGGCCAAGCCATTTATTAATCCAAAGACATGAC	6000
QY	6001	TCAAAATCAAAAGTGAAGAGAGATTAAGTGGAGATGGGGCTGTCAAGTGTGAGACAC	6060
Db	6001	TCAAAATCAAAAGTGAAGAGAGATTAAGTGGAGATGGGGCTGTCAAGTGTGAGACAC	6060
QY	6061	TGACCTTCACTTATTAAGTCACTAGGCGCAAGAGAGCTACAGAGGTTGACTGGTCTTA	6120
Db	6061	TGACCTTCACTTATTAAGTCACTAGGCGCAAGAGAGCTACAGAGGTTGACTGGTCTTA	6120
QY	6121	CTCAGCTTGGAGACAGGACGTGAGAAATGGGTGACTCCATTCCTATATGAAGAGGCTAG	6180
Db	6121	CTCAGCTTGGAGACAGGACGTGAGAAATGGGTGACTCCATTCCTATATGAAGAGGCTAG	6180
QY	6181	CACCAACAGGTACAGTGTCCCTGTGTCTCATGTCCAGAGATTCCTGGCCAGTTTCAAG	6240
Db	6181	CACCAACAGGTACAGTGTCCCTGTGTCTCATGTCCAGAGATTCCTGGCCAGTTTCAAG	6240
QY	6241	GACTTAAGACTCATCTGTGTGAAACAAAGTATCAAGCCCTAAGCCCATTTTGTGCT	6300
Db	6241	GACTTAAGACTCATCTGTGTGAAACAAAGTATCAAGCCCTAAGCCCATTTTGTGCT	6300
QY	6301	AATTAATCAGAACCCCTGGGGATGACGGCTCTGACAGCAGAGAGCTTTTAAAAAGCTC	6360

```
|||||
Db 6301 AATTAAATCAGAAACCCCTGGGATGCGAGCTCTGAGCAGCAGAGAGCTTTTAAAGAGCTC 6360
OY 6361 CCAGGTGATTCTGATCAGCAGCGTGAACAAACACACACTCAGGTTGCAACAGAAAGAGG 6420
Db 6361 CCAGGTGATTCTGATCAGCAGCTGGAACAAACACACACTCAGGTTGCAACAGAAAGAGG 6420
OY 6421 AAAGCTAGGGAAGGCTTGGGATGAGGAGGCTTCTTCAGGCGCAGTATGAGAGGCTGTT 6480
Db 6421 AAAGCTAGGGAAGGCTTGGGATGAGGAGGCTTCTTCAGGCGCAGTATGAGAGGCTGTT 6480
OY 6481 AGCAGTGGTGGCAGCTTCTCTCTGCTGTCATATAGCTATTCATCCATCCTATCCATC 6540
Db 6481 AGCAGTGGTGGCAGCTTCTCTCTGCTGTCATATAGCTATTCATCCATCCTATCCATC 6540
OY 6541 ACACCCACCCTCATCTTATGACCCGATCCCTTCATCCATCCATCCATCCATCCATCCAC 6600
Db 6541 ACACCCACCCTCATCTTATGACCCGATCCCTTCATCCATCCATCCATCCATCCATCCAC 6600
OY 6601 CCACGCATCCATCCAAACCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 6660
Db 6601 CCACGCATCCATCCAAACCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 6660
OY 6661 CATTTATCCACAGAGAACTGGTATTTGTAATATGAGGAGATTTAAATTTTAA 6720
Db 6661 CATTTATCCACAGAACTGGTATTTGTAATATGAGGAGATTTAAATTTTAA 6720
OY 6721 ACCTCTGTTGATTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 6780
Db 6721 ACCTCTGTTGATTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 6780
OY 6781 TGGCAATCGAGAAAGGTTTGGGTTGGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 6840
Db 6781 TGGCAATCGAGAAAGGTTTGGGTTGGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 6840
OY 6841 TTGAACTCAAAATTTATCGGCTGGTGGCAAGTCTTTACACCGACCAATTTTGGTGACA 6900
Db 6841 TTGAACTCAAAATTTATCGGCTGGTGGCAAGTCTTTACACCGACCAATTTTGGTGACA 6900
OY 6901 CATCATTTATTTAGAAACATCTTTATGATGATGATGATGATGATGATGATGATGATG 6960
Db 6901 CATCATTTATTTAGAAACATCTTTATGATGATGATGATGATGATGATGATGATGATG 6960
OY 6961 GCACGGATGACCTTTTAACTCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 7020
Db 6961 GCACGGATGACCTTTTAACTCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 7020
OY 7021 GTTCAACTGCTGTAATGCTTTTATTCACACACTCTGTGGGGGGGGGGGGGGGGGGATC 7080
Db 7021 GTTCAACTGCTGTAATGCTTTTATTCACACACTCTGTGGGGGGGGGGGGGGGGGGATC 7080
OY 7081 CCTGAGTTGGAGGCCGTTGGTCTGTCAGAGTTTCAGATACCTGGGGCTATACAGGAA 7140
Db 7081 CCTGAGTTGGAGGCCGTTGGTCTGTCAGAGTTTCAGATACCTGGGGCTATACAGGAA 7140
OY 7141 ACCCTATCCCAAAACAAACAAACAAACAAACAAACAAATTTCTGTGCATATATCACAGAG 7200
Db 7141 ACCCTATCCCAAAACAAACAAACAAACAAACAAACAAATTTCTGTGCATATATCACAGAG 7200
OY 7201 TTAGAGGATATTAAGTAGGAGTATGAGGCTGTGGAGGAGATCATGCTTTTGTATTT 7260
Db 7201 TTAGAGGATATTAAGTAGGAGTATGAGGCTGTGGAGGAGATCATGCTTTTGTATTT 7260
OY 7261 ATAAATAGTAAAGTACACAAAGATGATATATCTATCTATCTATCTATCTATCTATCTAT 7320
Db 7261 ATAAATAGTAAAGTACACAAAGATGATATATCTATCTATCTATCTATCTATCTATCTAT 7320
OY 7321 TATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTAT 7380
Db 7321 TATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTAT 7380
OY 7381 CTGCTTTGACTGTAATGCTCTATTTTCTGGGTCAACTCTTACCCCTAGTGTGGGTTT 7440
|||||

|||||
Db 7381 CTGCTTTGACTGTAATGCTCTATTTCTGGGTCAACTCTTACCCCTAGTGTGGGTTT 7440
OY 7441 ACCAACACCAGACATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 7500
Db 7441 ACCAACACCAGACATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 7500
OY 7501 GGACCTAGGAGTCTTGTGATGCTTAAGCAAGCTCTGCGACAGAGCTGCAGCTCCAGTCC 7560
Db 7501 GGACCTAGGAGTCTTGTGATGCTTAAGCAAGCTCTGCGACAGAGCTGCAGCTCCAGTCC 7560
OY 7561 CCATTTTGTTCAGGTACTCTGTGACAGTTGTATATTTGCGACCGCTATGTAAGTCTCTC 7620
Db 7561 CCATTTTGTTCAGGTACTCTGTGACAGTTGTATATTTGCGACCGCTATGTAAGTCTCTC 7620
OY 7621 CACCTCCAGTTCCACACTCTTGTGTCATCCAGGAGGGGGGCACTGTGCTCACA 7680
Db 7621 CACCTCCAGTTCCACACTCTTGTGTCATCCAGGAGGGGGGCACTGTGCTCACA 7680
OY 7681 GTGCCCCGTCTCCCTGCTTCAGACCTACATATTTGCTGTGACAGTTGATGTAATG 7740
Db 7681 GTGCCCCGTCTCCCTGCTTCAGACCTACATATTTGCTGTGACAGTTGATGTAATG 7740
OY 7741 GGATGCTGCTCTGTGATTTCTTTTATGCGTGGCCCTTTATCTTATGACAGATTTGTTG 7800
Db 7741 GGATGCTGCTCTGTGATTTCTTTTATGCGTGGCCCTTTATCTTATGACAGATTTGTTG 7800
OY 7801 GGCCATGTGCACTGCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTAT 7860
Db 7801 GGCCATGTGCACTGCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTAT 7860
OY 7861 TGTGTGATTAACCACTTCTGTTTCATTTACTGATGAAATTTGAGGCCACCCAC 7920
Db 7861 TGTGTGATTAACCACTTCTGTTTCATTTACTGATGAAATTTGAGGCCACCCAC 7920
OY 7921 CCTTTTTTTTTTATTTAGACAGAGCTTTTCTGTATATCTTTCATCTGCTGCTCCT 7980
Db 7921 CCTTTTTTTTTTATTTAGACAGAGCTTTTCTGTATATCTTTCATCTGCTGCTCCT 7980
OY 7981 GAGCTACTGCTGAGACAGAGGCTGTGAGGCTGCTCCACTTTTACACTCTGCTGAAAC 8040
Db 7981 GAGCTACTGCTGAGACAGAGGCTGTGAGGCTGCTCCACTTTTACACTCTGCTGAAAC 8040
OY 8041 AGAGTAGCCATGAACCTTCAAGACAAATTTCTGTTTGGTTTATTTTACATTTGTTG 8100
Db 8041 AGAGTAGCCATGAACCTTCAAGACAAATTTCTGTTTGGTTTATTTTACATTTGTTG 8100
OY 8101 TGTATGCTGTATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 8160
Db 8101 TGTATGCTGTATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 8160
OY 8161 TGGGACAGAGAAACCAAGATGCGCATTCCTCAGATACGATGCTGTTAAATATGTA 8220
Db 8161 TGGGACAGAGAAACCAAGATGCGCATTCCTCAGATACGATGCTGTTAAATATGTA 8220
OY 8221 TGTATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 8280
Db 8221 TGTATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 8280
OY 8281 CTTCCTTGTGGGCTCTCCGATTTAACTCAGCTCTCGGGGCTAGTGAAGCAATGCTTCA 8340
Db 8281 CTTCCTTGTGGGCTCTCCGATTTAACTCAGCTCTCGGGGCTAGTGAAGCAATGCTTCA 8340
OY 8341 CTGATGAGCAATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 8400
Db 8341 CTGATGAGCAATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 8400
OY 8401 CGCACTGCACTGGCTTAAAGCTCACCAAGTCACTCAGATGCGCTAGCAGGAGACTCAG 8460
Db 8401 CGCACTGCACTGGCTTAAAGCTCACCAAGTCACTCAGATGCGCTAGCAGGAGACTCAG 8460
OY 8461 GGATATGCTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 8520
Db 8461 GGATATGCTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 8520
```



Oy	8521	TTTTTAAACCGAATCCGAGAGATAGACAGGACTCACCAGTAAGGAGGGTCTTTTGT	8580
Db	8521	ATTTTAAACCTGAATCCCGAGGATAGAGACAGGACTCACCAGTAAGGAGGGTCTTTTGT	8580
Oy	8581	GTTTGTTGGTTCCTCTCGCATAGATCAGGACGTCGAATAGTGAAGCCTGGGCTAC	8640
Db	8581	GTTTGTTGGTTCCTCTCGCATAGATCAGGACGTCGAATAGTGAAGCCTGGGCTAC	8640
Oy	8641	ATACATCTCTCTCCAAAAGCCATAGTAGTAGGAGGCTGAGGCTTAACAAGCCT	8700
Db	8641	ATACATCTCTCTCCAAAAGCCATAGTAGTAGGAGGCTGAGGCTTAACAAGCCT	8700
Oy	8701	AAGCCGGCTGTGATAGCACACAGGATAGCCTGCACATATAGCAAGACCTTGTTCAAA	8760
Db	8701	AAGCCGGCTGTGATAGCACACAGGATAGCCTGCACATATAGCAAGACCTTGTTCAAA	8760
Oy	8761	ACATGAGAGGAGGGGATGTTTAAAGTCTGGGCTGTCTAACAGGACTTAAGGAGCCAA	8820
Db	8761	ACATGAGAGGAGGGGATGTTTAAAGTCTGGGCTGTCTAACAGGACTTAAGGAGCCAA	8820
Oy	8821	TGTAGACATTTGACATAGAAAGATCATCATCAAGCCGAGGGGACAGGTAGAGGTTGG	8880
Db	8821	TGTAGACATTTGACATAGAAAGATCATCATCAAGCCGAGGGTGGGACAGGTAGAGGTTGG	8880
Oy	8881	ACTACAGTGTCAAGACCCCATATAGGAAGCCAGTTTCCCTTCTCTGGGCTCAAGC	8940
Db	8881	ACTACAGTGTCAAGACCCCATATAGGAAGCCAGTTTCCCTTCTCTGGGCTCAAGC	8940
Oy	8941	CTGGCTGAGCGGACAGTCTCACAAGCCCTTCTCTAGGCTCGTCCACCAAG	8995
Db	8941	CTGGCTGAGCGGACAGTCTCACAAGCCCTTCTCTAGGCTCGTCCACCAAG	8995

RESULT 2	AC098570	192060 bp	DNA	linear	HTG 20-ANG-2002
LOCUS	AC098570	192060 bp	DNA	linear	HTG 20-ANG-2002
DEFINITION	Mus musculus clone RP23-278N11, WORKING DRAFT SEQUENCE, 6 unordered pieces.				
ACCESSION	AC098570				
VERSION	AC098570.2	GI:22325297			
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULFUP.				
SOURCE	Mus musculus (house mouse)				
ORGANISM	Mus musculus				
REFERENCE	Eunaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
AUTHORS	Birren, B., Nussbaum, C. and Lander, E.				
TITLE	Mus musculus, clone RP23-278N11				
JOURNAL	Unpublished				
REFERENCE	2 (bases 1 to 192060)				
AUTHORS	Birren, B., Linton, L., Nussbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S., Barna, N., Bastian, V., Boguslavsky, L., Boukhalter, B., Brown, A., Canarata, J., Campotano, A., Chang, J., Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A., Cooke, P., Dextrallano, K., Dewar, K., Diaz, J. S., Dodge, S., Faro, S., Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S., Gerde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N., Hagos, B., Heaford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kanat, A., Karatas, A., Kells, C., Laroque, K., Lamasares, R., Landers, T., Lehoczy, J., Levine, R., Liu, G., MacLean, C., Macdonald, P., Major, J., Margus, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., McPheters, R., Melrim, J., Meneus, L., Mhova, T., Mlenka, V., Murphy, T., Naylor, J., Nguyen, C., Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunhkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, W., Riley, R., Rise, C., Rogov, P., Roman, J., Roselli, M., Roy, A., Santos, R., Schauer, S., Schenck, R., Seaman, S., Severy, P., Spencer, B., Stange, R., Stojanovic, N., Straus, N., Subramanian, A., Talamas, J., Teeffaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Triggilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, Z., Zemdek, L., Zimmer, A. and Zody, M.				

TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 COMMENT  
 TITLE  
 JOURNAL  
 COMMENT  
 FEATURES  
 SOURCE

Direct Submission  
 Submitted (24-OCT-2001) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
 3 (bases 1 to 192060)  
 Birren, B., Nusbaum, C., Lander, E., All, A., Allen, N., Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguslavsky, L., Boukhaltier, B., Camarata, J., Chang, J., Chazaro, B., Choepel, Y., Collymore, A., Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J. S., Dodge, S., Faro, S., Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J., Geryna, S., Gord, S., Graham, L., Grand-Pierre, N., Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karates, A., Kells, C., Landers, T., Levine, R., Lindblad-Roh, K., Liu, G., Maclean, C., Macdonald, P., Major, J., Matthews, C., McCarthy, M., Meldrum, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norb, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunhkhang, P., Pierre, N., Raymond, C., Retta, R., Rise, C., Rogov, P., Roman, J., Roy, A., Schauer, S., Schnupback, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Talamas, J., Testaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.  
 Direct Submission  
 Submitted (20-AUG-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
 On Aug 20, 2002 this sequence version replaced g1:16356898.  
 All repeats were identified using RepeatMasker:  
 Smt, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>  
 Genome Center  
 Center: Whitehead Institute/ MIT Center for Genome Research  
 Center code: WtBR  
 Web site: <http://www-seq.wi.mit.edu>  
 Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)  
 Project Information  
 Center project name: L14796  
 Center clone name: 278\_N.11  
 Summary Statistics  
 Sequencing vector: Plasmid; n/a; 100% of reads  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Assembly program: Phrap; version 0.960731  
 Consensus quality: 190464 bases at least Q40  
 Consensus quality: 191061 bases at least Q30  
 Consensus quality: 191345 bases at least Q20  
 Insert size: 163000; agarose-1p  
 Insert size: 191560; sum-of-contigs  
 Quality coverage: 12.4 in Q20 bases; agarose-1p  
 Quality coverage: 10.5 in Q20 bases; sum-of-contigs  
 NOTE: This is a 'working draft' sequence. It currently consists of 6 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished number as soon as it is available and the accession number will be preserved.  
 1 4921: contig of 4921 bp in length  
 4922 5021: gap of 100 bp  
 5022 15200: contig of 10179 bp in length  
 15201 15300: gap of 100 bp  
 15301 78692: contig of 63392 bp in length  
 78693 78792: gap of 100 bp  
 78793 116118: contig of 37336 bp in length  
 116119 116218: gap of 100 bp  
 116219 174491: contig of 58273 bp in length  
 174492 174591: gap of 100 bp  
 174592 192060: contig of 17469 bp in length.  
 Location/Qualifiers  
 1. 192060  
 /organism="Mus musculus"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:10090"



```
misc_feature /clone-"RP23-278N11"  
1..4921 /clone_lib="RP23 Female Mouse BAC"  
/note="assembly_fragment  
clone_end:SP6  
vector_side:left"  
misc_feature 5022..15200 /note="assembly_fragment"  
15301..78692 /note="assembly_fragment"  
78793..116118 /note="assembly_fragment"  
116219..174491 /note="assembly_fragment"  
174592..192060 /note="assembly_fragment"  
misc_feature clone_end:T7  
vector_side:right"  
BASE COUNT 49668 a 46807 c 46516 g 48562 t 507 others  
ORIGIN
```

```
Query Match 97.4%; Score 8762.2; DB 2: Length 192060;  
Best Local Similarity 98.2%; Pred. No. 0;  
Matches 8962; Conservative 0; Mismatches 3; Indels 158; Gaps 6;
```

```
QY 1 GATCTGTCGACCAAGACACTGTGGTCCAGCACTGGGAGGTGGAGGAGGCTCA 60  
DB 40161 GATCTGTCGACCAAGACACTGTGGTCCAGCACTGGGAGGTGGAGGAGGCTCA 40220  
QY 61 GAAGTTAAAGTCATCCTTGTGTACATPAGCAAGGTTTCACGACCTTACATGAAA 120  
DB 40221 GAAGTTAAAGTCATCCTTGTGTACATPAGCAAGGTTTCACGACCTTACATGAAA 40280  
QY 121 CCTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 180  
DB 40281 CCTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 40340  
QY 181 CAGTGTGTCGACACACTTAAATTCAGATPAGCAAGGAGGAGGAGGAGGAGGAGGAGG 240  
DB 40341 CAGTGTGTCGACACACTTAAATTCAGATPAGCAAGGAGGAGGAGGAGGAGGAGGAGG 40400  
QY 241 GAGTTGAAAGTCACCTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 300  
DB 40401 GAGTTGAAAGTCACCTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 40460  
QY 301 AACCTGTCTCAATAAAGCAAGTAAGTACATPAGCAAGGAGGAGGAGGAGGAGGAGGAGG 360  
DB 40461 AACCTGTCTCAATAAAGCAAGTAAGTACATPAGCAAGGAGGAGGAGGAGGAGGAGGAGG 40520  
QY 361 GTTCATTTCAAGATGAGCACTCTATAGATGATTTCTTGAAGGAGGAGGAGGAGGAGGAGG 420  
DB 40521 GTTCATTTCAAGATGAGCACTCTATAGATGATTTCTTGAAGGAGGAGGAGGAGGAGGAGG 40580  
QY 421 TGGGGAAGGGAGTGGAGTCTCTAGATTAAAGTGTCTGAGGAGGAGGAGGAGGAGGAGGAGG 480  
DB 40581 TGGGGAAGGGAGTGGAGTCTCTAGATTAAAGTGTCTGAGGAGGAGGAGGAGGAGGAGGAGG 40640  
QY 481 TTGTGATTCATATGAAGGCTGTATAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 540  
DB 40641 TTGTGATTCATATGAAGGCTGTATAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 40700  
QY 541 GAAGAGGTGAGGAGGCTTATTAACATGAGCACTTATTAACATGAGCACTTATTAACATGAGCACT 600  
DB 40701 GAAGAGGTGAGGAGGCTTATTAACATGAGCACTTATTAACATGAGCACTTATTAACATGAGCACT 40760  
QY 601 CGTTCAAGTTGAAGATCACTTTCAACACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGG 660  
DB 40761 CGTTCAAGTTGAAGATCACTTTCAACACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGG 40820  
QY 661 GTAGGAGGAGGAGGAGGAGAGAGAGATTTAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGG 720  
DB 40821 GTAGGAGGAGGAGGAGGAGAGAGAGATTTAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGG 40880
```

```
QY 721 GCACATTATCATAACAGTGTCAAAATGTGCTAACTCCCTAGTCCACAGATGCTGTGTACA 780  
DB 40881 GCACATTATCATAACAGTGTCAAAATGTGCTAACTCCCTAGTCCACAGATGCTGTGTACA 40940  
QY 781 CTCGTTTCCTCTTCCATCTGTGTGACATTTGTGACAGACAGAGAAATAGAAATGTGGGT 840  
DB 40941 CTCGTTTCCTCTTCCATCTGTGTGACATTTGTGACAGACAGAGAAATAGAAATGTGGGT 41000  
QY 841 ATTTATTTGTGCTGTGAGACACATCCAGGAGGCTTTTACATTTACAGGACATGAGTTAC 900  
DB 41001 ATTTATTTGTGCTGTGAGACACATCCAGGAGGCTTTTACATTTACAGGACATGAGTTAC 41060  
QY 901 TAACTGGGCTACTTCTCCAAAGGTTTGAACCAATTTGTTTAAATTTACTTAAATTTGTGT 960  
DB 41061 TAACTGGGCTACTTCTCCAAAGGTTTGAACCAATTTGTTTAAATTTACTTAAATTTGTGT 41120  
QY 961 GCATGAGGATGAGCATGTATACGTATGTATGAGAGTCAATGATGCTGTACCTTCAAA 1020  
DB 41121 GCATGAGGATGAGCATGTATACGTATGTATGAGAGTCAATGATGCTGTACCTTCAAA 41180  
QY 1021 ATGATTCGATATCCCAAGAGTCAAGTCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1080  
DB 41181 ATGATTCGATATCCCAAGAGTCAAGTCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 41240  
QY 1081 GGAGCCAAAGGCTGGGTTCTGTGCAAGAGCAGCAGTGGCCTTAAACATGGGAGCAGCTCT 1140  
DB 41241 GGAGCCAAAGGCTGGGTTCTGTGCAAGAGCAGCAGTGGCCTTAAACATGGGAGCAGCTCT 41300  
QY 1141 CTAGGCTTAAGTATCTTTTATTTTAAATATATATTTCTACGCGGGGTGTGTGCT 1200  
DB 41301 CTAGGCTTAAGTATCTTTTATTTTAAATATATATTTCTACGCGGGGTGTGTGCT 41360  
QY 1201 ACAAGCCTTAAATCCCAAGCACTTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1260  
DB 41361 ACAAGCCTTAAATCCCAAGCACTTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 41420  
QY 1261 CTGGGGTGCAGAGCTTGCCCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1320  
DB 41421 CTGGGGTGCAGAGCTTGCCCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 41480  
QY 1321 CTTGCGGTATATGTCCTGTGCAAGGAGTGTGACATCCCTTGAAGTGTGAGTAAAGACAGTTG 1380  
DB 41481 CTTGCGGTATATGTCCTGTGCAAGGAGTGTGACATCCCTTGAAGTGTGAGTAAAGACAGTTG 41540  
QY 1381 TGATACAGGTCGCGTTACAGATGCTGGAATTTGAACCCAGGAGGAGGAGGAGGAGGAGGAGG 1440  
DB 41541 TGATACAGGTCGCGTTACAGATGCTGGAATTTGAACCCAGGAGGAGGAGGAGGAGGAGGAGG 41600  
QY 1441 AGTGTCTTAACTTGTGAGCCACCCCTCAACCCCTGCTTTAGAGACTTTAACTTTTG 1500  
DB 41601 AGTGTCTTAACTTGTGAGCCACCCCTCAACCCCTGCTTTAGAGACTTTAACTTTTG 41660  
QY 1501 TGTATGTGGAAGTGAAGTGTGACCTTACCAAGTGTGTGCTGCTGTAGCATCA 1560  
DB 41661 TGTATGTGGAAGTGAAGTGTGACCTTACCAAGTGTGTGCTGCTGTAGCATCA 41720  
QY 1561 CTGAGCCCGTACCAACAGCACTGATGATACAGTTTAAAGGCAACCTTAAACAATGACA 1620  
DB 41721 CTGAGCCCGTACCAACAGCACTGATGATACAGTTTAAAGGCAACCTTAAACAATGACA 41780  
QY 1621 ATAGTTGATAGAGTTTGAATATAGTCTAGAGTATTTGTTAGGAGTACCTTGTGCTGCC 1680  
DB 41781 ATAGTTGATAGAGTTTGAATATAGTCTAGAGTATTTGTTAGGAGTACCTTGTGCTGCC 41840  
QY 1681 TTAGCATGTGCTGTGAGAGATAGAAAATGAAAGACTTGTAGTGTCTGTGAACCCACA 1740  
DB 41841 TTAGCATGTGCTGTGAGAGATAGAAAATGAAAGACTTGTAGTGTCTGTGAACCCACA 41900  
QY 1741 GAGGAGGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1800  
DB 41901 GAGGAGGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 41960  
QY 1801 AATAGTTACAAATGATATAATAATTTAGTAAATTTCTTTTAAAGGTATATGTTGGAGGGA 1860
```

Db	41961	AATGTTACAAATGATAAATAATTAATTAGTAATTTCTTTTGAAGGATATATGTTGGAGGGA	42020
Oy	1861	GAGATGGCTCAGCTTCCAGAGACATTGCTCTCTTGAGAGAGACCTGATTCAGTTCC	1920
Db	42021	GAGATGGCTCAGCTTCCAGAGACATTGCTCTCTTGAGAGAGACCTGATTCAGTTCC	42080
Oy	1921	AGGACTCATATATGTTGGCTCACAGCCATCTGTAATCCAGTTCCAGAGGGTTCCACACCT	1980
Db	42081	AGGACTCATATATGTTGGCTCACAGCCATCTGTAATCCAGTTCCAGAGGGTTCCACACCT	42140
Oy	1981	CTTGCGGCTCCACAGGACACACATACATACACACATACATACATACAGCAAAACACC	2040
Db	42141	CTTGCGGCTCCACAGGACACACATACATACACACATACATACATACAGCAAAACACC	42200
Oy	2041	CATACACACATAAATTAATTAAGAAATTTAAAGGTGATGTTGTTGAACATTTGCT	2100
Db	42201	CATACACACATAAATTAATTAAGAAATTTAAAGGTGATGTTGTTGAACATTTGCT	42260
Oy	2101	TACACATGCTGATTGAGACATGTAACAAGCACACACTGAAGAGGGATCTGGGCTGGAG	2160
Db	42261	TACACATGCTGATTGAGACATGTAACAAGCACACACTGAAGAGGGATCTGGGCTGGAG	42320
Oy	2161	AGATGGCTCAGCGGTTAAGACACTGACTGCTTCGGAAGGAAGGTCTGAGTTCAAT	2220
Db	42321	AGATGGCTCAGCGGTTAAGACACTGACTGCTTCGGAAGGAAGGTCTGAGTTCAAT	42380
Oy	2221	CCTAGAACACACATGTTGGCTCACAAACCATCATATGAGATCTGACACCCCTCTTGCT	2280
Db	42381	CCTAGAACACACATGTTGGCTCACAAACCATCATATGAGATCTGACACCCCTCTTGCT	42440
Oy	2281	GCATCTGAAGACAGCTGACAGCTACAGTGTACTAGATATATCTAATTAATTAATCTTTT	2340
Db	42441	GCATCTGAAGACAGCTGACAGCTACAGTGTACTAGATATATCTAATTAATTAATCTTTT	42500
Oy	2341	TTTAAAAAATGAAAGAGGATCTGAGACCTCAAAAAGATTTATGACAGTGCACG	2400
Db	42501	TTTAAAAAATGAAAGAGGATCTGAGACCTCAAAAAGATTTATGACAGTGCACG	42560
Oy	2401	GGTATTAATCTATCTGTGAGTTTCTTTCCGCTTGCAACTGGGTGACAGACAG	2460
Db	42561	GGTATTAATCTATCTGTGAGTTTCTTTCCGCTTGCAACTGGGTGACAGACAG	42620
Oy	2461	CCCCCTTTTCAATCAAGAAAGGGGTGCTACATTAATTTCTGAACAAACACACACTGCAGT	2520
Db	42621	CCCCCTTTTCAATCAAGAAAGGGGTGCTACATTAATTTCTGAACAAACACACACTGCAGT	42680
Oy	2521	ATGTTTACTGCTGTTGCTGACTATGAGACAGCGACGGCGCGG--CACACACACAC	2578
Db	42681	ATGTTTACTGCTGTTGCTGACTATGAGACAGCGACGGCGCGGCGCGACACACACAC	42740
Oy	2579	ACACACACACACACACACACACACACACATTCAGTCTCCAGAGCTCTTGGGAAGT	2638
Db	42741	ACACACACACACACACACACACACACACATTCAGTCTCCAGAGCTCTTGGGAAGT	42800
Oy	2639	CAABAAGAGGCTGCCCTCAACACAGATCTTCTTCCCTCTTAAGAGAACACACAT	2698
Db	42801	CAABAAGAGGCTGCCCTCAACACAGATCTTCTTCCCTCTTAAGAGAACACACAT	42860
Oy	2699	CCAAGGTGGCAGAGATCTACAGGGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG	2758
Db	42861	CCAAGGTGGCAGAGATCTACAGGGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG	42920
Oy	2759	CCAGAGACCTACAGCAGAGGGCAGCAAGGAGATCCCAAGGTCCAGGGCAGGAGGTGA	2818
Db	42921	CCAGAGACCTACAGCAGAGGGCAGCAAGGAGATCCCAAGGTCCAGGGCAGGAGGTGA	42980
Oy	2819	GGCCCTTGTTCGAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG	2878
Db	42981	GGCCCTTGTTCGAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG	43040
Oy	2879	CAGCTCATAAAGTGGAGGTGCTGCTACATCAAAAGGAGGAGGAGGAGGAGGAGGAGG	2938
Db	43041	CAGCTCATAAAGTGGAGGTGCTGCTACATCAAAAGGAGGAGGAGGAGGAGGAGGAGG	43100
Oy	2939	GTGCCACATGAGCGAGGGTCTATGCTGATAGAGAGATCTGACAGGGTCCAGAGGCC	2998
Db	43101	GTGCCACATGAGCGAGGGTCTATGCTGATAGAGAGATCTGACAGGGTCCAGAGGCC	43160
Oy	2999	ACCTGTCTGTCCCAAGGGAGACCCCAAGTGTAACTGTGGCTTGGTGTCTAGTTCCAGC	3058
Db	43161	ACCTGTCTGTCCCAAGGGAGACCCCAAGTGTAACTGTGGCTTGGTGTCTAGTTCCAGC	43220
Oy	3059	TACAAGACCCAGGAGTCTCATCTCCATCCCATCCAGAGGCCCTCCGCCACACACC	3118
Db	43221	TACAAGACCCAGGAGTCTCATCTCCATCCCATCCAGAGGCCCTCCGCCACACACC	43280
Oy	3119	CACCCCGACATCCGTTGACCTTCTAGAGGCTGAGAGGGTGCACAGCCCTGTGGGGTT	3178
Db	43281	CACCCCGACATCCGTTGACCTTCTAGAGGCTGAGAGGGTGCACAGCCCTGTGGGGTT	43340
Oy	3179	GCCTACCTGCAAGTGAAGCCCAAGTCTTAGCCGGAAGTGCACCCATCTCTGAAGCTGCA	3238
Db	43341	GCCTACCTGCAAGTGAAGCCCAAGTCTTAGCCGGAAGTGCACCCATCTCTGAAGCTGCA	43400
Oy	3239	GAGCCAAAGGGGGGACACGCGACAGCTGAGGCTGACAGGCTGTTGGTGGCTAGGTTT	3298
Db	43401	GAGCCAAAGGGGGGACACGCGACAGCTGAGGCTGACAGGCTGTTGGTGGCTAGGTTT	43460
Oy	3299	CCAGGGACCTGGGACCTTCCACACCCCCATCCATCTCTCTGGGGCCCTATCTT	3358
Db	43461	CCAGGGACCTGGGACCTTCCACACCCCCATCCATCTCTCTGGGGCCCTATCTT	43520
Oy	3359	CCCTTATATGTTGAAGAAAGTTCTTGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG	3418
Db	43521	CCCTTATATGTTGAAGAAAGTTCTTGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG	43578
Oy	3419	CGGTCTCTGACGACGAGCTTCCACAACTTCTTAAGATCTCCAGTGGTGGCTGCTCT	3478
Db	43579	CGGTCTCTGACGACGAGCTTCCACAACTTCTTAAGATCTCCAGTGGTGGCTGCTCT	43638
Oy	3479	TCCAGACAGTGAAGCAATTTGGTGGGAGACACATGTTGACCAACAGTGTGGAGGGAC	3538
Db	43639	TCCAGACAGTGAAGCAATTTGGTGGGAGACACATGTTGACCAACAGTGTGGAGGGAC	43698
Oy	3539	AGGGTCTGCTGCTCTCTGACGCTGCTGCTTCTGAGACCTTGGTATTAAGTTGGG	3598
Db	43699	AGGGTCTGCTGCTCTCTGACGCTGCTGCTTCTGAGACCTTGGTATTAAGTTGGG	43758
Oy	3599	GGTAGGTGAAGTCTCTGAAACTCTGAAGAAGCAAGAACCCAGAGGCTGTCTTGGGC	3658
Db	43759	GGTAGGTGAAGTCTCTGAAACTCTGAAAGCAAGAACCCAGAGGCTGTCTTGGGC	43818
Oy	3659	CTTCAATGAAGAAAGTTTACAGACCCCTTCTGTAAGTCACTGCTGCTCATCTGTGT	3718
Db	43819	CTTCAATGAAGAAAGTTTACAGACCCCTTCTGTAAGTCACTGCTGCTCATCTGTGT	43878
Oy	3719	AGATTCCCTGGACCAAGGTGGCTCTGGGAGCTGAGATTTCTACAAATTAATGAGCA	3778
Db	43879	AGATTCCCTGGACCAAGGTGGCTCTGGGAGCTGAGATTTCTACAAATTAATGAGCA	43938
Oy	3779	GTCTGAGACTTGGACTCCGCTGTAATTAATTAATTAATTAATTAATTAATTAATTAAT	3838
Db	43939	GTCTGAGACTTGGACTCCGCTGTAATTAATTAATTAATTAATTAATTAATTAATTAAT	43998
Oy	3839	GTTCAATGCTTACACATCTGAATGATTTCTTGTGTCACATTCCTGACACTCTGG	3898
Db	43999	GTTCAATGCTTACACATCTGAATGATTTCTTGTGTCACATTCCTGACACTCTGG	44058
Oy	3899	GAGGTGCTATCTTGGACACATGATCTCGGGATGTAAGCTGACACACAGAGAGGG	3958
Db	44059	GAGGTGCTATCTTGGACACATGATCTCGGGATGTAAGCTGACACACAGAGAGGG	44118
Oy	3959	GAGAGTCAAGAGGTGTTCTTAGGCTTATTAAGGCTGACATCACCCCTTCTAGAA	4018
Db	44119	GAGAGTCAAGAGGTGTTCTTAGGCTTATTAAGGCTGACATCACCCCTTCTAGAA	44178



Db 46331 ACTGGGTCTACTACACTTGGAGCAGGACGTGGAATGGGACCTCATCTCATG 46390  
 QY 6170 AGAGGGCTGAGCAGCAGCAGGTACAGGTATCCCTGTCTCATGCGAGATTCGGCC 6229  
 Db 46391 AGAGGGCTGAGCAGCAGCAGGTACAGGTATCCCTGTCTCATGCGAGATTCGGCC 46450  
 QY 6230 AGTTTCAAGAGCAGTACAGTCTGTGTGTGAACAAGATCAAGCCCTAAGCC 6289  
 Db 46451 AGTTTCAAGAGCAGTACAGTCTGTGTGTGAACAAGATCAAGCCCTAAGCC 46510  
 QY 6290 CATTTTGTCTAATTAATCAGAACCCGGGGATCAGGCTGTGAGCAGGAGACTT 6349  
 Db 46511 CATTTTGTCTAATTAATCAGAACCCGGGGATCAGGCTGTGAGCAGGAGACTT 46570  
 QY 6350 TTAATAAGCTCCAGGTGATTTCTGATCAGCAGCTGGAACAAACAGCTACAGTTCA 6409  
 Db 46571 TTAATAAGCTCCAGGTGATTTCTGATCAGCAGCTGGAACAAACAGCTACAGTTCA 46630  
 QY 6410 CAGAAAGAGCAGAAAGCTAGGAAAGCTTGGGATGGGAGCCTTCCAGGCCAGTAGAT 6469  
 Db 46631 CAGAAAGAGCAGAAAGCTAGGAAAGCTTGGGATGGGAGCCTTCCAGGCCAGTAGAT 46690  
 QY 6470 GGAGGCTGGTTAGCAGTGGTGGGAGCTTCTGCTGCTGATAGCTATCCATCCACT 6529  
 Db 46691 GGAGGCTGGTTAGCAGTGGTGGGAGCTTCTGCTGCTGATAGCTATCCATCCACT 46750  
 QY 6530 CATCCATCCATACACCCACCCATCCATTTATGACACCATCTTCATCCATCCATCTATC 6589  
 Db 46751 CATCCATCCATACACCCACCCATCCATTTATGACACCATCTTCATCCATCCATCTATC 46810  
 QY 6590 CAGCTACCCAGCCAGCAGTCCATCCAAACCTCTCTTCTCTCTCTCTCTCTCTCTCT 6649  
 Db 46811 CAGCTACCCAGCCAGCAGTCCATCCAAACCTCTCTCTCTCTCTCTCTCTCTCTCTCT 46870  
 QY 6650 CTTCACCTCATCTTATTCACACAGAGAACTGGTATGTACTAAATGGGAGATTAAT 6709  
 Db 46871 CTTCACCTCATCTTATTCACACAGAGAACTGGTATGTACTAAATGGGAGATTAAT 46930  
 QY 6710 TAAATTTTAGAAGCTGTGTGATGTGACTATGTGCTATGTGAGCAGGTACATACCA 6769  
 Db 46931 TAAATTTTAGAAGCTGTGTGATGTGACTATGTGCTATGTGAGCAGGTACATACCA 46990  
 QY 6770 CAGCAGACGTGTGCAATGGGAAAGGTTTGGGTGTCTCTCTCTCTCTCTCTCTCTCT 6829  
 Db 46991 CAGCAGACGTGTGCAATGGGAAAGGTTTGGGTGTCTCTCTCTCTCTCTCTCTCTCT 47050  
 QY 6830 GGTTCGGGATTTGAATCAAAATTAATGGGCTGTGGCAAGTCTTTCCACAGGACCA 6889  
 Db 47051 GGTTCGGGATTTGAATCAAAATTAATGGGCTGTGGCAAGTCTTTCCACAGGACCA 47110  
 QY 6890 TTTTGTGACACATCATTTATTTATAGAAAGCATCTTATGTAGTCCAGGCTGGCCTACAG 6949  
 Db 47111 TTTTGTGACACATCATTTATTTATAGAAAGCATCTTATGTAGTCCAGGCTGGCCTACAG 47170  
 QY 6950 TTGCTATGTGCGCAGGATGACCTTTAACTCCTGCTCTTCCAGCCTCCACCGAGTGTCTA 7009  
 Db 47171 TTGCTATGTGCGCAGGATGACCTTTAACTCCTGCTCTTCCAGCCTCCACCGAGTGTCTA 47230  
 QY 7010 GGTTCAGGTGTCTAAC----- 7027  
 Db 47231 GGTTCAGGTGTCTAAC----- 47290  
 QY 7028 -----TGTGAAATGCTTTTAAATCCACAGCTCTGTGG 7060  
 Db 47291 AGAAATCTGGGCAACAGCTGGGCAATGGTGTGATGCTTTAAATCCACAGCTCTGTGG 47350  
 QY 7061 GGGGGGGGGAGAGGGGATCCCTGAGTTGAGGCCAGTTTGTCTACAGATTTTCAGAT 7120  
 Db 47351 GGGGGGGGGAGAGGGGATCCCTGAGTTGAGGCCAGTTTGTCTACAGATTTTCAGAT 47410  
 QY 7121 ACCTGGGCTATACAGGGAACCCCTATCCCAAAACAAACAAACAAACAAATATAT 7180  
 Db 7180 ACCTGGGCTATACAGGGAACCCCTATCCCAAAACAAACAAACAAACAAATATAT 7180

Db 47411 ACCTGGGCTATACAGGGAACCCCTATCCCAAAACAAACAAACAAACAAATATAT 47470  
 QY 7181 CTGTGCAATATACAGAGATTTAGAGATATTAGTAGGGCTGTGAGGAGA 7240  
 Db 47471 CTGTGCAATATACAGAGATTTAGAGATATTAGTAGGGCTGTGAGGAGA 47530  
 QY 7241 GTGATGCTTCTTGTGTATTAATTAATTAAGTACTACAGATGAT-----TATC 7292  
 Db 47531 GTGATGCTTCTTGTGTATTAATTAATTAAGTACTACAGATGAT-----TATC 47590  
 QY 7293 TATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATC 7352  
 Db 47591 TATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATC 47650  
 QY 7353 CATCCATCCATCATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATC 7412  
 Db 47651 CATCCATCCATCATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATC 47710  
 QY 7413 TCACACTTCCACCCCTAGTGTGGTTTACCAACACCCAGACATTTATTTATTTGTTT 7472  
 Db 47711 TCACACTTCCACCCCTAGTGTGGTTTACCAACACCCAGACATTTATTTATTTGTTT 47770  
 QY 7473 TATTTTATTAATCTAGAGCTCAGGGTGGGAGCTCAGGGTCTTGTGATGCTTAAGCAAGCT 7532  
 Db 47771 TATTTTATTAATCTAGAGCTCAGGGTGGGAGCTCAGGGTCTTGTGATGCTTAAGCAAGCT 47830  
 QY 7533 CTCTGCCACAGAGCTGACCTCCAGTCCCATTTTGTTCAGGTGACTCTGTGACAGTTGT 7592  
 Db 47831 CTCTGCCACAGAGCTGACCTCCAGTCCCATTTTGTTCAGGTGACTCTGTGACAGTTGT 47890  
 QY 7593 CATATTCGAGGGCTATGTAGCTCTCTCCACCTCCAGTTCAGACCTTCTGTGATCTC 7652  
 Db 47891 CATATTCGAGGGCTATGTAGCTCTCTCCACCTCCAGTTCAGACCTTCTGTGATCTC 47950  
 QY 7653 CAGTGGGGGCAACTCTGTGCTCACCAGTGGCCCTTCCCTGCTTCCAGACCTACATAT 7712  
 Db 47951 CAGTGGGGGCAACTCTGTGCTCACCAGTGGCCCTTCCCTGCTTCCAGACCTACATAT 48010  
 QY 7713 TTGCTGTCTGAACAGTTGATGTAATGGGATGCTTCTGTGTATCTTTATAGGCTGG 7772  
 Db 48011 TTGCTGTCTGAACAGTTGATGTAATGGGATGCTTCTGTGTATCTTTATAGGCTGG 48070  
 QY 7773 CCCCTTATCTAGCAGACTTGTGTGGGCAATGTGTACAGTCTATCTATCTATCTATC 7832  
 Db 48071 CCCCTTATCTAGCAGACTTGTGTGGGCAATGTGTACAGTCTATCTATCTATCTATCTATC 48130  
 QY 7833 ATCATCTTAATGCTTAATGCTTCTTGTGTGATTAACACATCTTCTGTTCATTTTAC 7892  
 Db 48131 ATCATCTTAATGCTTAATGCTTCTTGTGTGATTAACACATCTTCTGTTCATTTTAC 48190  
 QY 7893 TGAATGAAATTTGTGGCCGCCACCCACCTTTTATTTATTTGAGACAAGGCTTTC 7952  
 Db 48191 TGAATGAAATTTGTGGCCGCCACCCACCTTTTATTTATTTGAGACAAGGCTTTC 48250  
 QY 7953 TGTGTATCTTGAATCTTGGCTGTCTGAGTCTCATCTGTGTAGACAGGCTGTGAGGCTG 8012  
 Db 48251 TGTGTATCTTGAATCTTGGCTGTCTGAGTCTCATCTGTGTAGACAGGCTGTGAGGCTG 48310  
 QY 8013 TCCCTTCACCTTTTGAACCTCTGTGAAAGAGTAGCATGAACCTTCAAGACAAATTTTCT 8072  
 Db 48311 TCCCTTCACCTTTTGAACCTCTGTGAAAGAGTAGCATGAACCTTCAAGACAAATTTTCT 48370  
 QY 8073 GTTTTGGTTTGTTTTATCATTTTGTGTGTATAGCTGTATATGTGACATGTTGTGCTT 8132  
 Db 48371 GTTTTGGTTTGTTTTATCATTTTGTGTGTATAGCTGTATATGTGACATGTTGTGCTT 48430  
 QY 8133 CAGGTGCTACATGTGTGACTGTGTGTGGCAGAGAAACAAACGATGTGCAATCTCT 8192  
 Db 48431 CAGGTGCTACATGTGTGACTGTGTGTGGCAGAGAAACAAACGATGTGCAATCTCT 48490  
 QY 8193 CAGATCTACAGATCTGTGTATATGTATGTATGTATGTATGTATGTATGTATGTATGTAT 8252  
 Db 48491 CAGATCTACAGATCTGTGTATATGTATGTATGTATGTATGTATGTATGTATGTATGTATGTAT 48550

```

OY 8253 TATGAGGTATTTTGGAGTTTTCACCTTCCCTGTGGGCTCCGATTAACCTCAG 8312
    |||||||
Db 48551 TATGAGGTATTTTGGAGTTTTCACCTTCCCTGTGGGCTCCGATTAACCTCAG 48610
OY 8313 CTCTCGGGGTAGTGAAGCAATGCTTCACTCGATGAGCCATCTCGCTCCCTGTCGCA 8372
    |||||||
Db 48611 CTCTCGGGGTAGTGAAGCAATGCTTCACTCGATGAGCCATCTCGCTCCCTGTCGCA 48670
OY 8373 CCTCTCTTATTTTCCAGATGAGCACTGACATGCGCTTAAGCTCAACCAAGTCA 8432
    |||||||
Db 48671 CCTCTCTTATTTTCCAGATGAGCACTGACATGCGCTTAAGCTCAACCAAGTCA 48730
OY 8433 TCCAGAGTGGTACGACGAGGACACTCAGGATATGCGCTTCGCTCCAGAGTACG 8492
    |||||||
Db 48731 TCCAGAGTGGTACGACGAGGACACTCAGGATATGCGCTTCGCTCCAGAGTACG 48790
OY 8493 AATTACAGGATACATCATCTGCTGAGAGATTTTAACTGAATCCTGAGATAGAGCAG 8552
    |||||||
Db 48791 AATTACAGGATACATCATCTGCTGAGAGATTTTAACTGAATCCTGAGATAGAGCAG 48850
OY 8553 CACTCTACCATGAGGGCTTCTTTTGTGTTGGTTTCCCTGCTCAATAGATCAGG 8612
    |||||||
Db 48851 CACTCTACCATGAGGGCTTCTTTTGTGTTGGTTTCCCTGCTCAATAGATCAGG 48910
OY 8613 CAGTCTGAATAGTGTAGCTGCGCTACATTAACATCTGTCTCAAAAAGCCTATAGAGT 8672
    |||||||
Db 48911 CAGTCTGAATAGTGTAGCTGCGCTACATTAACATCTGTCTCAAAAAGCCTATAGAGT 48970
OY 8673 AGGAGAGTGAAGGCTAAAGAAAGAGCCTTAAGCCGGCTGTGATAGCACAGAGATAGCCTG 8732
    |||||||
Db 48971 AGGAGAGTGAAGGCTAAAGAAAGAGCCTTAAGCCGGCTGTGATAGCACAGAGATAGCCTG 49030
OY 8733 CACATATATGCAAGACCTTCTTCAAAAACATGAGGAGGCTATGTTTAAAGCTCGG 8792
    |||||||
Db 49031 CACATATATGCAAGACCTTCTTCAAAAACATGAGGAGGCTATGTTTAAAGCTCGG 49090
OY 8793 GCTGTGTAAACAGGCACTAAGGAGGACCAATGTAGACATTTGACTAAGAAAGATCATCATC 8852
    |||||||
Db 49091 GCTGTGTAAACAGGCACTAAGGAGGACCAATGTAGACATTTGACTAAGAAAGATCATCATC 49150
OY 8853 AAAGCCGGGTGGGCGAGGTAGAGTGTGACTACAGTGTCAAGACCCCATAGAGAGCCA 8912
    |||||||
Db 49151 AAAGCCGGGTGGGCGAGGTAGAGTGTGACTACAGTGTCAAGACCCCATAGAGAGCCA 49210
OY 8913 GTTTCCTCTTCTCCGCTGAGGCTCAAGCCTGAGGCGACGCTCAACATGAGCCTT 8972
    |||||||
Db 49211 GTTTCCTCTTCTCCGCTGAGGCTCAAGCCTGAGGCGACGCTCAACATGAGCCTT 49270
OY 8973 CTCTCTAGGCTGTCCACCATG 8995
    |||||||
Db 49271 CTCTCTAGGCTGTCCACCATG 49293

```

```

RESULT 3
AC132885
LOCUS AC132885 170074 bp DNA linear HTG 12-MAR-2003
DEFINITION Mus musculus clone RP24-263015, WORKING DRAFT SEQUENCE, 9 unordered
          pieces
ACCESSION AC132885
VERSION AC132885.3 GI:28927760
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 170074)
AUTHORS Birren,B., Nusbaum,C., and Lander,E.
TITLE Mus musculus, clone RP24-263015
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 170074)
AUTHORS Birren,B., Nusbaum,C., Lander,E., All,A., Allen,N., Anderson,S.,
          Batra,N., Bastien,V., Bloom,T., Boguslavsky,L., Boukhgalter,B.,

```

TITLE  
JOURNAL  
REFERENCE  
AUTHORS

TITLE  
JOURNAL  
COMMENT

Camata,J., Chang,J., Chazaro,B., Choepel,Y., Collamore,A.,  
Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S.,  
Faro,S., Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J.,  
Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,  
Horion,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kanat,A.,  
Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,  
Liu,G., Maclean,C., Macdonald,P., Major,J., Matthews,C.,  
McCarthy,M., Meldrum,J., Menus,L., Mihova,T., Mlengv,V.,  
Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,  
O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K.,  
Phunkhang,P., Plerie,N., Raymond,C., Retla,R., Rise,C., Rogov,P.,  
Roman,J., Roy,A., Schauer,S., Schupbach,R., Seaman,S., Severy,P.,  
Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,  
Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,  
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,  
Zembek,L., Zimmer,A. and Zody,M.

Direct Submission  
Submitted (04-SEP-2002) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
3 (bases 1 to 170074)

Birren,B., Nusbaum,C., Lander,E., Abouelkell,A., Allen,N.,  
Anderson,S., Arachchi,H.M., Batra,N., Bastien,V., Bloom,T.,  
Boguslavsky,L., Boukhgalter,B., Camata,J., Chang,J., Choepel,Y.,  
Collamore,A., Cooke,P., Cook,A., Cooke,P., Corum,B., Dearellano,K.,  
Diaz,J.S., Dodge,S., Dooley,K., Dorris,B., Erickson,J., Faro,S.,  
Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J., Gardyna,S.,  
Graham,L., Grand-Pierre,N., Hages,N., Hagopian,D., Hagos,B.,  
Hall,J., Horion,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,  
Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,  
Lindblad-Toh,K., Liu,G., Lui,A., Mabbitt,R., Maclean,C.,  
Macdonald,P., Major,J., Manning,D., Matthews,C., McCarthy,M.,  
Meldrum,J., Menus,L., Mihova,T., Mlengv,V., Murphy,T., Naylor,J.,  
Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,  
O'Neill,D., Oliver,J., Peterson,K., Phunkhang,P., Plerie,N.,  
Rachupka,A., Ramasamy,V., Raymond,C., Retla,R., Rise,C., Rogov,P.,  
Roman,J., Schauer,S., Schupbach,R., Seaman,S., Severy,P., Smith,C.,  
Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,  
Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,  
Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,  
Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

Direct Submission  
Submitted (12-MAR-2003) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Mar 12, 2003 this sequence version replaced g1:28416187.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: MIBR  
Web site: http://www-seq.wi.mit.edu  
Contact: sequence\_submissions@genome.wi.mit.edu

----- Project Information  
Center project name: L27142  
Center clone name: 263.O\_15  
----- Summary Statistics  
Sequencing vector: Plasmid; n/a; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.960731  
Consensus quality: 168429 bases at least Q40  
Consensus quality: 168822 bases at least Q30  
Consensus quality: 169112 bases at least Q20  
Insert size: 164000; agarose-fp  
Insert size: 169274; sum-of-ctnfigs  
Quality coverage: 8.7 in Q20 bases; agarose-fp  
Quality coverage: 8.4 in Q20 bases; sum-of-ctnfigs

----- NOTE: This is a 'working draft' sequence. It currently  
\* consists of 9 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence





4093 AGTTTGGACTACTCAGACCAAGTTTATCTGCTGAGAAATCCCTGTGACATTGAGTT 4152  
1379 AGTTTGGACTACTCAGACCAAGTTTATCTGCTGAGAAATCCCTGTGACATTGAGTT 4138  
4153 GGGAGATTTGCTCTGGGGGCTTCCAGGTCTTTGGTTAGCAGAGGGATTCCTTTGTATA 4212  
1439 GGGAGATTTGCTCTGGGGGCTTCCAGGTCTTTGGTTAGCAGAGGGATTCCTTTGTATA 4198  
4213 GGGCATGACCTAGTGTCTATGTTTACTAATTCCTGTCCAGTTTAAAGCTGGAACCTAAA 4272  
1499 GGGCATGACCTAGTGTCTATGTTTACTAATTCCTGTCCAGTTTAAAGCTGGAACCTAAA 1558  
4273 CCCAGGCGAGCCGCCAGATTCTCTACAGTTGTACCCCAAGAACACAGACAGTATGATA 4332  
1559 CCCAGGCGAGCCGCCAGATTCTCTACAGTTGTACCCCAAGAACACAGACAGTATGATA 1618  
4333 TGCAGAGATAGTACGTGGGAGAGAGAACTTAAACCCGCCCAAGGCCACAGAGTTTC 4392  
1619 TGCAGAGATAGTACGTGGGAGAGAGAACTTAAACCCGCCCAAGGCCACAGAGTTTC 1678  
4393 GCTTCCCTAGTTCACAATGCCAGTATGAGTGTAGTATGAGTGGGTGTGAGTTGGTAGC 4452  
1679 GCTTCCCTAGTTCACAATGCCAGTATGAGTGTAGTATGAGTGGGTGTGAGTTGGTAGC 1738  
4453 TACAAGCATGATGATGTTTCATGTGTGTATGTTATCTGAGCACTTGGAGAGCTGA 4512  
1739 TACAAGCATGATGATGTTTCATGTGTGTATGTTATCTGAGCACTTGGAGAGCTGA 1798  
4513 AGCAGAGATTTGCTATATGTTTGAAGGCCAGCCTGAGCTATAGAGCGAGACTTGTCTTT 4572  
1799 AGCAGAGATTTGCTATATGTTTGAAGGCCAGCCTGAGCTATAGAGCGAGACTTGTCTTT 1858  
4573 AAGAAAAAATGAAAGCCCGAGAGTGTGGCAGACGCCCTTAATCCACACTTGGGAGG 4632  
1859 AAGAAAAAATGAAAGCCCGAGAGTGTGGCAGACGCCCTTAATCCACACTTGGGAGG 1918  
4633 CAGAAGCAGGAGATTTTCTGAGTTTCAAGGCCAGCCTGTCTATAGAGTATCCAGAGC 4692  
1919 CAGAAGCAGGAGATTTTCTGAGTTTCAAGGCCAGCCTGTCTATAGAGTATCCAGAGC 1978  
4693 AGCCAGGGCTACACAGAGAAACCCCTGTTTGAAGAAACGAGAAACCAAAACAAA 4752  
1979 AGCCAGGGCTACACAGAGAAACCCCTGTTTGAAGAAACGAGAAACCAAAACAAA 2038  
4753 CAAAACAAAACCCAAACCCAAACCCCTCATCTCTCATCTCTCTAGGCTGTCT 4812  
2039 CAAAACAAAACCCAAACCCAAACCCCTCATCTCTCTCATCTCTCTAGGCTGTCT 2098  
4813 GTCTAGGTGTAGAGTTTGGGAGCTTCAGACTTATATATATAATAGCCCTTTTATCACTG 4872  
2099 GTCTAGGTGTAGAGTTTGGGAGCTTCAGACTTATATATAATAGCCCTTTTATCACTG 2158  
4873 GTCAAGAGCAGAGAAAGTTTCACTGTGGGACACAGTGGGACCTTGAGAAAGTACTCTTGG 4932  
2159 GTCAAGAGCAGAGAAAGTTTCACTGTGGGACACAGTGGGACCTTGAGAAAGTACTCTTGG 2218  
4933 CCAGCCCAAAATTTCTGGGAAGGCTTCTGGAGGAAGTGTGCCGA----- 4979  
2219 CCAGCCCAAAATTTCTGGGAAGGCTTCTGGAGGAAGTGTGCCGATTGAGTCTGGGAAG 2278  
4980 -----TCAG 4983  
2279 AATATGAAGAGTGGGAAAGCCCAAGGCAAGAGCAGCCTGGGTGTTAAAGCAGCAGCTCAG 2338  
4984 ACTACTGTTTCTGAAGGAGAGAGAGAGGTTGGAAGATGTTGGTGGACAGACAGTTGGA 5043  
2339 ACTACTGTTTCTGAAGGAGAGAGAGAGGTTGGAAGATGTTGGTGGACAGACAGTTGGA 2398  
5044 ACAGAGAGCAGAGAGGAGGAGCATTCCAAGATTTCTGAGACATGATGACTGCTTTGGTTCT 5103  
2399 ACAGAGAGCAGAGAGGAGGAGGAGCATTCCAAGATTTCTGAGACATGATGACTGCTTTGGTTCT 2458

5104 CTGGGTGACAAAGTGTCCCGAGGATAGGGCTGTAGAAAGGGGACCGGGGTGAGCCAAAT 5163  
2459 CTGGGTGACAAAGTGTCCCGAGGATAGGGCTGTAGAAAGGGGACCGGGGTGAGCCAAAT 2518  
5164 GAGTTCAAGTTGAGGAGACATCCAGAGGCCCTTCTGTGGCAAGCTTAAGAAATGAGA 5223  
2519 GAGTTCAAGTTGAGGAGACATCCAGAGGCCCTTCTGTGGCAAGCTTAAGAAATGAGA 2578  
5224 GCCCTTAACCCCTCCCTGAAAGTTTAGGGAGACAGAGAGCTGAGAGATTCCTTCTAGGG 5283  
2579 GCCCTTAACCCCTCCCTGAAAGTTTAGGGAGAGAGAGAGCTGAGAGATTCCTTCTAGGG 2638  
5284 TGAAGAGAGATATGCTGTCTGACCAACATGGCTTAGAGACAGAGAGTGGACCGTTA 5343  
2639 TGAAGAGAGATATGCTGTCTGACCAACATGGCTTAGAGACAGAGAGTGGACCGTTA 2698  
5344 CCCCTCAGAACCAAGCCATCCCTCTTGGCTCTAAGGAGCTGGGCCCTTTCTGTTTAAAG 5403  
2699 CCCCTCAGAACCAAGCCATCCCTCTTGGCTCTAAGGAGCTGGGCCCTTTCTGTTTAAAG 2758  
5404 AATCTTACTTTCTTCTCAGAGAGAGCAGAGAGCCTTGTCCCTCCCTGTTGGTCAATA 5463  
2759 AATCTTACTTTCTTCTCAGAGAGAGCAGAGAGCCTTGTCCCTCCCTGTTGGTCAATA 2818  
5464 ACACCCCTGTGTAAACATTAGTTATTTTACTGTGCTAGTTTGTCTCAGAGACAGTCCATCT 5523  
2819 ACACCCCTGTGTAAACATTAGTTATTTTACTGTGCTAGTTTGTCTCAGAGACAGTCCATCT 2878  
5524 GGTAGACCTGTCTCTTAACATCAACCAAGGTATGGCCCAATTCCTCACCAGAAAGTGC 5583  
2879 GGTAGACCTGTCTCTTAACATCAACCAAGGTATGGCCCAATTCCTCACCAGAAAGTGC 2938  
5584 AGAAGAGACCTTGAAGAAAGGTTACATTAACAAGATGGCAGATAAACAATAACT 5643  
2939 AGAAGAGACCTTGAAGAAAGGTTACATTAACAAGATGGCAGATAAACAATAACT 2998  
5644 ACTATCTTCTTGTACCCAAATTTGTTTGTCTGAACAGAGAGGGGTGTGAGTGTATGTG 5703  
2999 ACTATCTTCTTGTACCCAAATTTGTTTGTCTGAACAGAGAGGGGTGTGAGTGTATGTG --- 3054  
5704 TGT 5763  
3055 ---TGT 3110  
5764 GGGGACCTTTTATGCTTAAGAAATATCTATTTTGGCGCCATGCAACAGGGGTATTGG 5823  
3111 GGGGACCTTTTATGCTTAAGAAATATCTATTTTGGCGCCATGCAACAGGGGTATTGG 3170  
5824 GGAGAGTCAAGCTTCTGCAAAACACAGTAAGTGTCCCAAGATGATGTGGCTGAAATCA 5883  
3171 GGAGAGTCAAGCTTCTGCAAAACACAGTAAGTGTCCCAAGATGATGTGGCTGAAATCA 3230  
5884 CCAAGGGGAGGCTGATCAGAGTGGACAGACATCTCAAGATTAAGCCACCTGTGGGCT 5943  
3231 CCAAGGGGAGGCTGATCAGAGTGGACAGACATCTCAAGATTAAGCCACCTGTGGGCT 3290  
5944 CAGAAGAGGAGTTTACAGAGGTAAAGGCCAAGCCATTATTATCAAGACATGACCTCA 6003  
3291 CAGAAGAGGAGTTTACAGAGGTAAAGGCCAAGCCATTATTATCAAGACATGACTCA 3350  
6004 AATCAAAATGCAAGAGAGATTTAGCTGAGAGATGGGCTGTCACTGTGGGACACCTGA 6063  
3351 AATCAAAATGCAAGAGAGATTTAGCTGAGAGATGGGCTGTCACTGTGGGACACCTGA 3410  
6064 CCTTGACATTATTAGTCACTAGGCCAAGAGAGATCAGAGAGGTAAGTGGTCTTACTC 6123  
3411 CCTTGACATTATTAGTCACTAGGCCAAGAGAGATCAGAGAGGTAAGTGGTCTTACTC 3470  
6124 AGCTTGGAGCAAGGACGTGGAGAAATGGGAGCTTCATCTGATGAGAGGGGCTCAGCAC 6183  
3471 AGCTTGGAGCAAGGACGTGGAGAAATGGGAGCTTCATCTGATGAGAGGGGCTCAGCAC 3530  
6184 CACCAAGTACAGTGTTCCTGTCTGTCTATGCCAGAGATTCTGTGGCAGTTTCAAGAGAC 6243

|||||  
Db 3531 CACCAGGTACAAAGTTCCTGTGTCTCAGCCAGGATTCCTGGCCAGTTTCAAGGAC 3590  
OY 6244 TAAGAGCTCATCTGTGTGGAACAAGATATCCAAAGCCCTAAGCCCATTTGGTCTAAT 6303  
Db 3591 TAAGAGCTCATCTGTGTGGAACAAGATATCCAAAGCCCTAAGCCCATTTGGTCTAAT 3650  
OY 6304 TAAATCAGAAACCCCTGGGGATGCAAGGCTGTGACAGCAGAGAGCTTTTAAAAAGCTCCA 6363  
Db 3651 TAAATCAGAAACCCCTGGGGATGCAAGGCTGTGACAGCAGAGAGCTTTTAAAAAGCTCCA 3710  
OY 6364 GGTGATTCGTATCAGCAGCTGGAACAACACAGTACAGGTTCAACAGAAAGAGGAAA 6423  
Db 3711 GGTGATTCGTATCAGCAGCTGGAACAACACAGTACAGGTTCAACAGAAAGAGGAAA 3770  
OY 6424 GCTAAGGAAAGCTTTGGGATGGGAGCCCTTCTCAGAGCCAGTATGAGAGCTGTATAGC 6483  
Db 3771 GCTAAGGAAAGCTTTGGGATGGGAGCCCTTCTCAGAGCCAGTATGAGAGCTGTATAGC 3830  
OY 6484 AGTGTGGCAGGCTTCTCTGCTGTCAATATAGCTATCCATCCATCATCCATCATCA 6543  
Db 3831 AGTGTGGCAGGCTTCTCTGCTGTCAATATAGCTATCCATCCATCATCCATCATCA 3890  
OY 6544 CCCAGCCATCCATTTATGACCCATCCATCCATCCATCCATCCATCCATCCATCCATCC 6603  
Db 3891 CCCAGCCATCCATTTATGACCCATCCATCCATCCATCCATCCATCCATCCATCCATCC 3950  
OY 6604 CGCATCCATCCAAACCTTCTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCAT 6663  
Db 3951 CGCATCCATCCAAACCTTCTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCAT 4010  
OY 6664 TTATCCACAGAGAACTGGTATGATGTAATATGGAGATTTAATTTTATAGAGC 6723  
Db 4011 TTATCCACAGAGAACTGGTATGATGTAATATGGAGATTTAATTTTATAGAGC 4070  
OY 6724 TCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 6783  
Db 4071 TCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 4130  
OY 6784 CAATCGAGAAAGGTTTTGGGTGTGTTTTCTCTCCACCGTGGGTTCTGGGATG 6843  
Db 4131 CAATCGAGAAAGGTTTTGGGTGTGTTTTCTCTCCACCGTGGGTTCTGGGATG 4190  
OY 6844 AACTCAAAATATGGGCTGTGTGCAAGTCTTTACACAGCCATTTTGTGACACAT 6903  
Db 4191 AACTCAAAATATGGGCTGTGTGCAAGTCTTTACACAGCCATTTTGTGACACAT 4250  
OY 6904 CATTATTTATAGAAAGCATTTATGATGATGATGATGATGATGATGATGATGATG 6963  
Db 4251 CATTATTTATAGAAAGCATTTATGATGATGATGATGATGATGATGATGATGATG 4310  
OY 6964 CGGATGACCTTTAAGTCTGCTCTCTCCAGCCTCACCCGAGTCTAGGTTTACAGGTGT 7023  
Db 4311 CGGATGACCTTTAAGTCTGCTCTCTCCAGCCTCACCCGAGTCTAGGTTTACAGGTGT 4370  
OY 7024 CAAC----- 7027  
Db 4371 CAACACACACCAAAATTTAAGTCTCAGGATTTTGTACTGAGCAGGCAAAATCTGGGCAA 4430  
OY 7028 -----TGGTAATGCTTTAATCCAGACCTGTGTGGGGGGGGGGAGG 7074  
Db 4431 CAGCTGGGCAATGTTGGGATGCTTTAATCCAGACCTGTGTGGGGGGGGGGAGG 4490  
OY 7075 CGGATCCCTGATTTGGAGGCAAGTGTGCTACAGAGTTTCAAGATTAACCTGGGGCTATAC 7134  
Db 4491 CGGATCCCTGATTTGGAGGCAAGTGTGCTACAGAGTTTCAAGATTAACCTGGGGCTATAC 4550  
OY 7135 AGGGAACCTTAATCCAAACAAACAAACAAACAAACAAATTTCTGTGCAATATCA 7194  
Db 4551 AGGGAACCTTAATCCAAACAAACAAACAAACAAACAAATTTCTGTGCAATATCA 4610  
OY 7195 CAGAGATTAGAGATTTATAGGATGATGAGGCTGGTGGAGAGATCATGCTTCTTT 7254  
|||||

Db 4611 CAGAGATTAGAGATTTATAGTACGGCTGTGAGGAGAGATCATGCTTCTTT 4670  
OY 7255 TGTATTAATAATAGTAAGTACTCACAAGATGAT-----TATCATATCTATCTA 7306  
Db 4671 TGTATTAATAATAGTAAGTACTCACAAGATGATTTATCTATCTATCTATCTA 4730  
OY 7307 TCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTA 7366  
Db 4731 TCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTA 4790  
OY 7367 TCGATTAAGCCAGGCTGCTTGAATGAGTCCATTTCTGGGCACTCTCACCC 7426  
Db 4791 TCGATTAAGCCAGGCTGCTTGAATGAGTCCATTTCTGGGCACTCTCACCC 4850  
OY 7427 CTAGTGTGGGTTTACCAACACCAGACATTTATTTATTTTATTTATTTATCT 7486  
Db 4851 CTAGTGTGGGTTTACCAACACCAGACATTTATTTATTTTATTTATTTATCT 4910  
OY 7487 AGGAGCTCAGGGTGGGACTCAGGCTCTTGATGCTAAGCAAGCTCTGCCACAGAGC 7546  
Db 4911 AGGAGCTCAGGGTGGGACTCAGGCTCTTGATGCTAAGCAAGCTCTGCCACAGAGC 4970  
OY 7547 TGCAGCTCCAGTCCCATTTTGTTCAGGTGACTCTGAGACAGTTGCAATTTGCGAGCG 7606  
Db 4971 TGCAGCTCCAGTCCCATTTTGTTCAGGTGACTCTGAGACAGTTGCAATTTGCGAGCG 5030  
OY 7607 TATGAGCTCTCTCAACCTCCAGTTCACAGACTTCTGTGCTATCCAGTGGGCGCA 7666  
Db 5031 TATGAGCTCTCTCAACCTCCAGTTCACAGACTTCTGTGCTATCCAGTGGGCGCA 5090  
OY 7667 CTCTGTGCTCACCAGTGGCTGTCCCTGTCTCTCAGACCTACATATTTGGCTGTGAA 7726  
Db 5091 CTCTGTGCTCACCAGTGGCTGTCCCTGTCTCTCAGACCTACATATTTGGCTGTGAA 5150  
OY 7727 AGTTCAATGTAATGGGATGGCTCTGCTGATTTCTTTATGGCTGGCCCTTATCTTAG 7786  
Db 5151 AGTTCAATGTAATGGGATGGCTCTGCTGATTTCTTTATGGCTGGCCCTTATCTTAG 5210  
OY 7787 CACAGTTTGTGTTGGGCATGTGCTACGTCTATCTATCTATCTATCTATCTATGCT 7846  
Db 5211 CACAGTTTGTGTTGGGCATGTGCTACGTCTATCTATCTATCTATCTATCTATGCT 5270  
OY 7847 TAATAGTGTCTCTTGTGTGATTAACCACTTCTGTTTCAATTTACTGATGGAATTTGT 7906  
Db 5271 TAATAGTGTCTCTTGTGTGATTAACCACTTCTGTTTCAATTTACTGATGGAATTTGT 5330  
OY 7907 GGCCCAACCCCAACCTTTTCTTTTATTTGAGACAAGGCTCTTCTGTGTAATCTTGA 7966  
Db 5331 GGCCCAACCCCAACCTTTTCTTTTATTTGAGACAAGGCTCTTCTGTGTAATCTTGA 5390  
OY 7967 ATCTGGCTGTCTGAGCTCAGTCTGTAGACAAGGCTGTGAGGCTGTCTCCACTTTTG 8026  
Db 5391 ATCTGGCTGTCTGAGCTCAGTCTGTAGACAAGGCTGTGAGGCTGTCTCCACTTTTG 5450  
OY 8027 ACACTCTGTGAACAAGTACGATTAACCTTCAAAACAAATTTCTGTGTTTGGTGTGT 8086  
Db 5451 ACACTCTGTGAACAAGTACGATTAACCTTCAAAACAAATTTCTGTGTTTGGTGTGT 5510  
OY 8087 TTTTCAATTTGTGTGTATGCGTGTATATGATGATTTGTGTCTCAGGTGCTCAATG 8146  
Db 5511 TTTTCAATTTGTGTGTATGCGTGTATATGATGATTTGTGTCTCAGGTGCTCAATG 5570  
OY 8147 TGTGTACCTGTGTGTGGAGACAGAAACAAACGATGTGCCATTTCTCAGATCTACGAT 8206  
Db 5571 TGTGTACCTGTGTGTGGAGACAGAAACAAACGATGTGCCATTTCTCAGATCTACGAT 5630  
OY 8207 CTGTGTAATATGATATGATATGATTTATTTAGTGTGCCAAGTATGACGATATTT 8266  
Db 5631 CTGTGTAATATGATATGATATGATTTATTTAGTGTGCCAAGTATGACGATATTT 8326  
OY 8267 GTTGGAGTTTTCACCTTCCCTTGTGGGCTCTCCGATTTAAACCTCAGCTCCGCGGTAG 8326  
Db 5691 GTTGGAGTTTTCACCTTCCCTTGTGGGCTCTCCGATTTAAACCTCAGCTCCGCGGTAG 5750



OY		8327	GAGCAATGCCCTTCACTCGATGATGGACCATTCGTGCCTCCCTGCTGCACACTCCTCCTTAATT	8386
Db		5751	GAGCAATGCCCTTCACTCGATGATGGACCATTCGTGCCTCCCTGCTGCACACTCCTCCTTAATT	5810
OY		8387	CCCGAATGGGACTACGCGACTGCTACTGGCGCTTAAACCTCACCAAGTCATCCAGATGGCTAG	8446
Db		5811	CCCGAATGGGACTACGCGACTGCTACTGGCGCTTAAACCTCACCAAGTCATCCAGATGGCTAG	5870
OY		8447	CCAGGGAGACTCAGGGGATATGCTGGCCTCTGCTCCACAGTGTAGAATTACAGGCATAC	8506
Db		5871	CCAGGGAGACTCAGGGGATATGCTGGCCTCTGCTCCACAGTGTAGAATTACAGGCATAC	5930
OY		8507	ATCATGCTGGAAAGTTTTTTAACCTGMAATCCTGGAGATAGGCGAGGCACTCTACCATAAG	8566
Db		5931	ATCATGCTGGAAAGTTTTTTAACCTGMAATCCTGGAGATAGGCGAGGCACTCTACCATAAG	5990
OY		8567	AGGGTTCTTTTGTTGGTTGGTTGGTTCCCTCTCATAAAGATCAGGCAAGTCTGAATAAGT	8626
Db		5991	AGGGTTCTTTTGTTGGTTGGTTGGTTCCCTCTCATAAAGATCAGGCAAGTCTGAATAAGT	6050
OY		8627	GTACCCCTGGGCTAACAATAACATCTTGCTCCAAAAAGCCTATPAGAGTAGGAGAGTGCAGGC	8686
Db		6051	GTACCCCTGGGCTAACAATAACATCTTGCTCCAAAAAGCCTATPAGAGTAGGAGAGTGCAGGC	6110
OY		8687	TAAAGAAGAGCCTTAAGCGCGCTGTGATAGACACAGAGATPAGCCTGGACCTTATATAGAAC	8746
Db		6111	TAAAGAAGAGCCTTAAGCGCGCTGTGATAGACACAGAGATPAGCCTGGACCTTATATAGAAC	6170
OY		8747	ACCTTGTTTCAAAAACATGAGGAGAGGGGATATGTTTTAAAGTCTGGGCTGTGTAAACAGC	8806
Db		6171	ACCTTGTTTCAAAAACATGAGGAGAGGGGATATGTTTTAAAGTCTGGGCTGTGTAAACAGC	6230
OY		8807	ACTAAGGAGGCCATGTATGACATTTTGGACTAAGAAGATTCATCATCAAAGCCGGGTGGGC	8866
Db		6231	ACTAAGGAGGCCATGTATGACATTTTGGACTAAGAAGATTCATCATCAAAGCCGGGTGGGC	6290
OY		8867	AGGGTAGAGTGTTGGACTACAGTGGTCAAGACCCCCTATAGAAAGCCAGTTTCCCTTCTTCC	8926
Db		6291	AGGGTAGAGTGTTGGACTACAGTGGTCAAGACCCCCTATAGAAAGCCAGTTTCCCTTCTTCC	6350
OY		8927	TCTGGGCCCTCAGCCCTGGCTCGACGGCCATGCTCTCACATGCTCTTCTCTTAGGCTCG	8986
Db		6351	TCTGGGCCCTCAGCCCTGGCTCGACGGCCATGCTCTCACATGCTCTTCTCTTAGGCTCG	6410
OY		8987	TCCACCATG 8995	
Db		6411	TCCACCATG 6419	
RESULT 4 LOCUS AC115173				
DEFINITION	Rattus norvegicus clone CH230-286017, WORKING DRAFT SEQUENCE, 3			
ACCESSION	AC115173			
VERSION	AC115173.4 GI:25006753			
KEYWORDS	HMG: HMG; PHASE1: HMG; DRAFT: HMG; FULLTOP.			
SOURCE	Rattus norvegicus (Norway rat)			
ORGANISM	Rattus norvegicus			
REFERENCE	1 (bases 1 to 193472)			
AUTHORS	Muzny,D.,Marle,M.,Metzker,M.,Lee,A.,Abramson,S.,Adams,C.,Alder,J.,Allen,C.,Allen,H.,Alsbrooks,S.,Amis,A.,Angiano,D.,Anyalebechi,V.,Aoyagi,A.,Ayodeji,M.,Baca,E.,Baden,H.,Baldwin,D.,Bandaranaikie,D.,Barber,M.,Barnstead,M.,Benahmed,F.,Blawie,K.,Blair,J.,Blankenburg,K.,Blyth,P.,Brown,M.,Bryant,N.,Buch,C.,Burch,P.,Buttelli,K.,Calderon,E.,Cardenas,V.,Carter,K.,Cavazos,I.,Ceasar,H.,Center,A.,Chacko,J.,Chavez,D.,Chen,G.,Chen,R.,Chen,Y.,Chen,Z.,Chu,J.,Cleaveland,C.,Cockrell,R.,Cox,C.,Coyle,M.,Gree,A.,D'Souza,L.,			

TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 COMMENT

Davila,M.L., Davis,C., Day-Carroll,L., De Anda,C., Dederich,D.,  
 Delgado,O., Denson,S., Derramo,C., Ding,Y., Dinh,H., Divya,K.,  
 Draper,H., Dugan-Rocha,S., Dunn,A., Dublin,K., Duval,B., Eaves,K.,  
 Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,  
 Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,  
 Frisner,C.M., Gabisi,A., Gante,R., Garcia,A., Garner,T., Garza,M.,  
 Gergeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,T., Guevara,W.,  
 Gunaratne,P., Haaland,W., Hamli,C., Hamilton,C., Hamilton,K.,  
 Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,  
 Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,  
 Hollins,B., Howells,S., Huliy,S., Hume,J., Idelbroid,D., Jackson,A.,  
 Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,  
 Kapachy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,  
 Kwis,C., Kraft,C.L., Lebow,H., Levan,D., Lewis,L., Li,Z., Liu,J.,  
 Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,  
 Lorenshaw,L., Loulseged,H., Lozada,R.J., Lu,X., Ma,J.,  
 Maheshwari,M., Mahindratne,M., Mahmood,M., Malloy,K., Mangum,A.,  
 Mangun,B., Mapa,P., Martin,K., Martin,R., Martinez,E.,  
 Manthey,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,  
 Mollasavilevic,A., Miner,G., Ming,E., Montemayor,J., Moore,S.,  
 Morgan,M., Morris,K., Morris,S., Mundasa,M., Murphy,M., Naif,L.,  
 Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,  
 Nwokwelench,O., Okwunou,G., Olarunsaogun,A., Pal,S., Parks,K.,  
 Pasternak,S., Paul,H., Perez,A., Perez,L., Pinnkoch,C.,  
 Plopper,F., Polndexter,A., Popovic,D., Primus,E., Pu,L.,L.,  
 Plazzo,M., Quirzo,J., Rachlin,E., Reeves,K., Regier,M., Reigh,R.,  
 Riley,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,  
 Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,  
 Sanders,W., Savary,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,  
 Shetty,J., Shwartsbeyn,A., Slasson,I., Sitter,C.D., Smajs,D.,  
 Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,  
 Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,  
 Taylor,T., Thomas,N., Thomas,S., Tinley,A., Trejos,Z., Usmani,K.,  
 Valas,R., Vanga,Y., Villasana,D., Waldron,L., Walker,B., Wang,J.,  
 Wang,O., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,  
 Williams,G., Willson,R., Wleczky,R., Wooden,H., Worley,K.,  
 Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,  
 Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von  
 Niederhausen,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,  
 Weinstock,G. and Gibbs,R.A.

Direct Submission  
 Unpublished  
 2 (bases 1 to 193472)  
 Worley,K.C.  
 Direct Submission  
 Submitted (15-MAR-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 3 (bases 1 to 193472)  
 Rat Genome Sequencing Consortium.  
 Direct Submission  
 Submitted (15-NOV-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 On Nov 15, 2002 this sequence version replaced gi:23678661.  
 The sequence in this assembly is a combination of BAC based reads  
 and whole genome shotgun sequencing reads assembled using Atlas  
 (<http://www.hgsc.bcm.tmc.edu/projects/atlas/>). Each contig described  
 in the feature table below represents a scaffold in the Atlas  
 assembly (a 'contig-scaffold'). Within each contig-scaffold,  
 individual sequence contigs are ordered and oriented, and separated  
 by sized gaps filled with Ns to the estimated size. The sequence  
 may extend beyond the ends of the clone and there may be sequence  
 contigs within a contig-scaffold that consist entirely of whole  
 genome shotgun sequence reads. Both end sequences and whole genome  
 shotgun sequence only contigs will be indicated in the feature  
 table.

----- Genome Center  
 Center: Baylor College of Medicine  
 Center code: BCM  
 Web site: <http://www.hgsc.bcm.tmc.edu/>  
 Contact: hgsc-help@bcm.tmc.edu  
 Project Information



OY	1272	ACCTTGCGCCCTGTTTTTTTTTTGTTTTTTCTTTATATGACACTGGTGTCTTACCGCGTAT	1331
Db	120742	AAGAAAAAATATATATATATCTCTGTGCATCTGTGACATGCTATATTTCCCGAC	120801
OY	1332	GT-----	1333
Db	120802	TTGAAGAGGTGGGGGTGGGGATCATAAATTTCACTGGGGTAAAGATGTGGCCCTGTT	120861
OY	1334	-----CGGTGCAGGGGTGCAGATCCCTTGGAGGTGAGTAAAGACAGTTGTGATCAC	1387
Db	120862	TTTCCCCCTTATATGTCGACCTGTGTCTTGCCCTGTGTGTGTAGGGAATCCGATCCCTG	120921
OY	1388	GCTGCCGTTTACAGATGCTGGAAATTAAGACCCAGGTGTCCCTAGAGAGACGCCAGTGTCTC	1447
Db	120922	GAGCTCCGATGGGTGCTGGAATTAAGACCCAGACACCCCTGGAAAGAGCCCAATGCTTC	120981
OY	1448	TTAACTTCGAGCCACCCCTCCAACTGCTTTTATAGACTCTTAACCTTTTGTGTAAAG	1507
Db	120982	TTAACTTCGAGCCACCCCTCCAGCCCTTAAGAAATTCCTAAACCTTTTGTGTGTGT	121041
OY	1508	TGGGAACAGAT-----	1544
Db	121042	GTGTGTGTGTGTGTGAATATGGAATTTGAACCTTAGAGGCTTGTACACTAGTGTGTG	121101
OY	1545	CTGCGCTGTAGCATCTAGAGCCGCTACCCACAGCACTAGTGATACAGTTTAAAGGCAA	1604
Db	121102	CTGTCTGAATATCATCTAGGCCACACCCACAGCTGATGATGATATATCTAATGCGAA	121161
OY	1605	ACACTTAACATGACAAATAGTTGGATAGATGATTGAAT--TATCTGTAGCTATTTGGTTA	1662
Db	121162	ACATTATACAAAGACAAAGATTTTGGGTAGTGTGAAGACCCAGTCTGTGACTATTTGGTTA	121221
OY	1663	GGGTACCTTTGCTGCTTACATATGTGTGTGAGAAATGAAAAATGAAAGCTTGAGT	1722
Db	121222	GGGTACCTTTGCTGCTTACATATGTGTGTGAGAAATGAAAAATGAAAGCTTGAGT	121275
OY	1723	CTAATCTGGAACCCACAGAGGCGAGGCAACCACTCTGAAAGTTTGTCTGTAGCT	1782
Db	121276	TGAGTCCCGGACATGCAATGTGTGGGTGGAACCCACCTCTGAAGTTG--TGTGTAGCT	121334
OY	1783	TCACATACACATTCACATTAATAGTTACATGATTAATTAATTAATTAATTTCTTTTAAA	1842
Db	121335	CTGCATACAAAC---CGTAAAGATTATATATGACAAATGAATTAAGAGAGCTTTTATGAA	121390
OY	1843	GGTATATGTTGGGAGGAGAGATGGTGCCTTCCAGAGAGCACTTGCCTGTCGAGAG	1902
Db	121391	GGTATATGTTGGGAGGAGAGATGGTGCCTTCCAGAGAGCACTTGCCTGTCGAGAG	121446
OY	1903	GACCTAGATTCAATTCACAGACTATATGTGGCTCACAGGCATCTGTAAATCCAGTTC	1962
Db	121447	GCCCTAGATTCAATTCACAGACTATATGTGGCTCACAGGCATCTGTAAATCCAGTTC	121505
OY	1963	CAGAGGTTCCACACCCCTTCTTGCGCTCCACAGGCACCACTATCTAGTACACAGACAT	2022
Db	121506	CAGAGGATTTTACACCCCTTCTTGCGCTCCACAGGCACCACTATCTAGTACAAACACAT	121565
OY	2023	ACATGCAGGCAAAACACCATACACATTAATTAATTAAGAAACCTTAAAGTGATGT	2082
Db	121566	ACATGCAGGCG--AAACCCCATATACACATTAATTAATTAAGAAACCTTAAAGTGATGT	121624
OY	2083	GTGTGTAAACATTTGTCTTACACATGCTGATTAAGAGACATGTACAAACGACACAGTGAAG	2142
Db	121625	GTGTG-----TGTGCATACACATGTGTCTGTAAAGACATGTGTGCAACACACACACA--	121675
OY	2143	AGGATCTGGGGCTGGAGATGGCTTAAGAGCACTGACTGCTCTTCCGAAG	2202
Db	121676	-----	121675
OY	2203	AAGTCTGAGTTTCAATTCATGCAACCACTGTGTGCTCAACAACTCATATATGAGAT	2262
Db	121676	-----CACACACACACACA-----	121691
OY	2263	CTGACACCCCTTTGTGTGATCTGAAAGACAGCTGACAGCTACAGTGTACTTAAATATA	2322

Db	121692	-----	121691
Oy	2323	CTAATAATTAATCTTTTAAAAAATGAGAGGATCTGAGACACCTCAAAAGAT	2382
Db	121692	-----CACACACTGGGAAGAGGCTCAGACACTTCAMAAATAGAT	121730
Oy	2383	TATGAGCAGTACTACACGGGTATTTATCTCTGAGCTTTTTCGGCTGGCTTG	2442
Db	121731	TATGGGCACTGACTCATGGGTGACTATCTATCTGAGTCTTTTTCCTTGGCTTGGCTTG	121790
Oy	2443	CACTGGGTGGACAGACGCCCTTTTCACTTACAAAGAACGGGTGCTACATTATTTCTGAA	2502
Db	121791	TAACTGGGTGGACAGGCACTCTTTCATTCACAAAGATGAGTGTACTATTATTTCTTAA	121850
Oy	2503	CAAAACACACCTTGACATGATGTTTACTCTCTGCTACTATGAGACAGCCAGCCGCGC	2562
Db	121851	CAAAACACACCTTGACATGATGTTTACTCTCTGCTACTATGAGACAGCCAGCCGCGC	121907
Oy	2563	GCGCACACACACACACACACACACACACACACACACACATTCAGTCTCCA	2622
Db	121908	-----CACACACACTTTCAGTCTCCA	121930
Oy	2623	GAGCTTTGGGAAGGTCAAGAAAGGCTGCCCTCAAAACAGATCTTCACTTTCCCTCT	2682
Db	121931	GAGCTTTTGGGAAGGTCAAGGAAGGCGCCCTCAAAACCTCTTCACTTCCCTCTCT	121990
Oy	2683	AAAGGAGACACGATTTCCAGGTGGCAGAAATCTACAGGGGGCAGAGGCAGGAGGGG	2742
Db	121991	AA--GGAACCAAGATTTCTAAGTGGCAGAAATCTTACAGGGGGCAGAGGCAGGAGGGG	122048
Oy	2743	AAGCAGCCATGTTTCCAGAGACTACAGACAGAGGGCAGCAGAGCAGATCCCAAGTCC	2802
Db	122049	AAGCAGGACACAGTTTCCAGAGACTACAGCAGAGGGCAGCAGACAGATCCCAAGTCC	122108
Oy	2803	AGGGAGGAGGAGTGGAGGCCCTTGTTCCGAGGA-----GAAGGACGGGGGAGAAACGGG	2857
Db	122109	CGGGAGGAGGAGAGAGGCCCTTGCTCTCTGGAGAAAGAGGTTCCGGAGAGCAGGG	122168
Oy	2858	TTCAAAAGCA--CAGTTTATGGCAGACTCATAAAGTGGAGGTCTGGCTCACTCAGAAAG	2916
Db	122169	TTCAAAAGCATTTGGTTATGGCAACCTATAAAGGGAGTTGGG---CTCAGAAAG	122224
Oy	2917	GAGCAAGAAAGGAAAGGCCCTTGTGCCACTGAGCGAGGGTCAATGCTGATGAGAGAT	2976
Db	122225	GAGGAGGACTGGAAAGGTCTTGTGCCACTGAGCGAGGGTCAATGCTGATGAGAGAT	122284
Oy	2977	CTGCAGGGGTCCACAGAGACCCCACTGCTCTCCCAAGGAAACCCCAAGTGTGAACCTG	3036
Db	122285	CTGAAGAGGTGCCAGGGGCCCACTGT-----CCCAAGGAAACCCCAAGTGTGAACCTG	122340
Oy	3037	GCTTGTGGTGTGATGTTCCAGCTACAAAGACCCAGGAGTCTCTACTCCATCCCACT	3096
Db	122341	GCTTGTGGTGTGATGTTCCAGCTGCAAGGCCCAAGAGTCTCTGTCNNNNNNNNNNNN	122400
Oy	3097	GCCCCCTGCCCCGCCACA-----CCCCACCCCGCACTCCCGTGC	3136
Db	122401	NN	122460
Oy	3137	CACATTCTTGAAGCTGAGC-----	3156
Db	122461	TACTTTTCTTAGGCGAGAGGGTGTGCTCTCAGATTCCCCCATCTTTGGCGGGAAGAGAA	122520
Oy	3157	-----GTGGCCAGCCCTGTGTGGGGGTTGGCTTACTCTGAGTGTAGAGCCCAAGT	3203
Db	122521	TGTCGCCGACACTGTGGCTCTCCCTGGGGGAGATGTCTACTCTCAGGGAAGAGTCTAGT	122580
Oy	3204	CTTAGCGGAGTGCACCCCATCTCCCTGAGCTGAGAGGCCAAGGGGGGGGCGACAGCGCAG	3263
Db	122581	CTTAGGGCGAAGTGCACCCCATCTCCCTGAGCTGAGAGGCC--AGGGCGGAGACAGCGCAG	122639
Oy	3264	CTCAGGCTGTCAGGCTGTTGCTGGGCTTACGTTCCAGAGGACCTGGGCACTTACTTCCC	3323





D 126726 TGGGGCTGGAGGAGGAGGAGGATGCTTTCTTTGTATTATATCTATCTGGATTATCTAT 126785  
Q 7243 CAGCTTTCTTTTGTATTAATAGTAAGTA----- 7274  
D 126786 CTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCT 126845  
Q 7275 ----- 7274  
D 126846 TCCATTTCTGTGTACAGCTTCTCCAGTGTAGGGTTACAGGCAATTCATCTATCATC 126905  
Q 7275 CTCACAGATGATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCT 7332  
D 126906 CAGACATTTTATTTCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTAT 126965  
Q 7333 -----TACCTACCTACCTACCTATC----- 7352  
D 126966 CTATCTATCTAACATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCT 127025  
Q 7353 -----CATCCATCTCATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTAT 7408  
D 127026 CTATCTATCTATCTAACATCTATCTATCTATCTATCTATCTATCTATCTATCTATCT 127085  
Q 7409 TGGGTCAACTCTTCACCCCTAGTGGTGGTTTACCAACACCCAGACATTTAT----- 7460  
D 127086 TATCTATCTAACATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCT 127145  
Q 7461 -----TTTTTTTGTATTTATTTATTAATCT 7486  
D 127146 TATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCT 127205  
Q 7487 AGGAGCTCAGGGGAGGAGCTGCTTGGCANTGCTAAGCAAGCTCTGCCACAGAGC 7546  
D 127206 AGAGATTAAAGGGGGAGCTGGGGCCCTGTGTATGTAAACAAGCTCTGTGCACAGAGC 127265  
Q 7547 TGCAGCTCAGTCCCATTTTGTTCAGGTGACTGTGAGACATGTGATCTGCAACGC 7606  
D 127266 TCCAGCCCAAGTCCCATCTTG-TCAGGTATCTGTGGGCTTGACATTTTCACAATGC 127324  
Q 7607 TATGAGCTCTCTCCACCTCCAGTTCACAGCTTCTGTCTATCCAGTGGCGGGCAA 7666  
D 127325 TATGAGCTCTCTCCACCTCTAGTTCACAGCTTCTGTCTATCCAGT-----GGGCAA 127380  
Q 7667 CTCTGAGCTCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 7703  
D 127381 CTCTGAGCTTACAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 127440  
Q 7704 CCTACATATTTGCTGTCTGAAACAGTTCATGTAATGAGATGAGTCTGTCTATCTTT 7763  
D 127441 CCTACAGATTTGCTGTCTGAAACAGTTCATGTAATGAGATGAGTCTGTCTGTCTTT 127500  
Q 7764 TATGAGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 7823  
D 127501 TATGAGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 127554  
Q 7824 TATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCT 7883  
D 127555 CATTTT-----ATCATCTATGAGCTTAATAGTGTCTGTGTGTGTGTGTGTGTGTGTGT 127611  
Q 7884 TTTCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTAT 7943  
D 127612 TTTCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCT 127671  
Q 7944 AGGTCTTTCTGTGTAACT-TGCAATCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 8002  
D 127672 TATTTTGTAGGCTCTTTCTATGATGATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 127731  
Q 8003 TGTGAGGCTGTCTCTTCCACTTTTGACATCTGTGTGAACAGAGATGACATGATCTCAAG 8062  
D 127732 TGTG-----GTTCACCTTTTGACATCTGTGTGAACAGAGATGACATGATCTCAAG 127783  
Q 8063 ACAATTTCTGT 8122  
D 127784 ACACTTTCT-----AGTTTTTCTTATTTTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 127832

Q 8123 TTTGTCTCTCAGTGTCTCATATGTGTGTACCTGTGTGT-----GGACAGAGAAACAAACCG 8179  
D 127833 TTTGTGTCTAGTGTCTCATATGT 127892  
Q 8180 ATGTGCCATTCCTCAGTACTAGGATCTGTGTATATGTATGTATGTATGTATGTATGTATGT 8239  
D 127893 ATGTGCCATTCCTCAGTACTAGGATCTGTGTATATGTATGTATGTATGTATGTATGTATGT 127940  
Q 8240 TAGTGTCCCAAGATATCAGATATTTTGTGTGAGTTTTCACCTTCCCTTGTGGGCTCTCC 8299  
D 127941 TAGTGTCCCA-----GTGCATTTAG 127963  
Q 8300 GCATTTAACTCAGCTCTGTGGGCTAGTGAACATGCTTGCATCTGATGACCATCTGTCT 8359  
D 127964 GGAATTGAATCTGGCTCTCAGGTTTACCGA-----GCCCTCAGTACCTGTCCCTTGTCTG 128018  
Q 8360 GCCCTGTCTGCCACCTCTCTCTTATTTTCCAGATGGAGACTACGACCTGTGCTCTTAA 8419  
D 128019 CCCCTGTCTGCCACCTCTCTCTTATTTTCCCA-----GACAGGACTCTACCTGTGCTCTTAA 128075  
Q 8420 GCTCACCAGTATCTCAGATGTGCTAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 8479  
D 128076 GCTCACCAGGACATCCGAGCTGTGCTAGCTGTGGGAGACCCAGGATATGCTGTCTTACC 128135  
Q 8480 TCCAGAGTCTAGATTTACAGGATATCATCTGTCTGTGAGATTTTAACTGAAATCTG 8539  
D 128136 TCCATGATGTAGATCTCAAGATATCAGTACGTT-----TGAAATCTG 128180  
Q 8540 AGATATGAGAGGACCTCTACCAATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 8584  
D 128181 AAGATATGAGAGGACCTCTACCAATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 128240  
Q 8585 GGTGTGTCTCTCTGTATGAATCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 8634  
D 128241 GGTGTGTCTCTCTGTATGAATCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 128300  
Q 8635 GGTCTATTAACATCTGTCTCAAAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 8686  
D 128301 GGTCTATTAACATCTGTCTCAAAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 128360  
Q 8687 TAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 8746  
D 128361 TAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 128420  
Q 8747 ACCTTGTCTCAAAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 8806  
D 128421 ACCTTGTCTCAAAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 128480  
Q 8807 ACTAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 8866  
D 128481 ACTAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 128540  
Q 8867 AGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 8925  
D 128541 AGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 128600  
Q 8926 CTCTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 8954  
D 128601 CTCTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 128660  
Q 8955 ACTGCTCTCAGATCT 8995  
D 128661 ATTTGCTCTCAGATCT 128701  
RESULT 5  
AC101018 149109 bp DNA linear HTG 21-AUG-2002  
LOCUS Rattus norvegicus clone RP32-328P7, \*\*\* SEQUENCING IN PROGRESS:\*\*\*  
DEFINITION 9 unordered pieces.  
AC101018  
AC101018 2 GT:22381320  
VERSION









Db 98979 GAGCTTTTGGGAAGGTCAAGGGAGGCTGCCCTCAACACCCTCTTCATCTTCCCTCCT 99038  
Oy 2683 AAAGGAGACACGATTTCCAAAGGTGGCAGAAATCTACAGGGGGCAGAGGAGGAGGGG 2742  
Db 99039 AA--GGAAACAAGATTTCTAAGGTGGCAGAAATTTCTACAGGGGCCAGAGGAGGAGGGG 99096  
Oy 2743 AAGCAGGCGATGTTTCCAGAGACCTACAGCAGAGGGGAGCAAGGCAGATCCCAAGTCC 2802  
Db 99097 AAGCAGGACACAGTTTCCAGAACCTACAGCAGAGGGGAGCAAGGCAGATGCCCAAGTCC 99156  
Oy 2803 AGGGCAGGAGGTGGAGGCCCTTGTTCGAGG----AGAAGCAGCGGGCAGAAACAGG 2857  
Db 99157 CGGGAGGGAGAGGAGGCCCTTGGCTCTCGGAGAAAGGAGTTCCGGCAGACAGG 99216  
Oy 2858 TTCAAAGCA--CAGTTTATGGCAGCTCATAAAGTGGAGTGGTGGGCTCACTCAGAAAG 2916  
Db 99217 TTCAAAGCATTTGGTTTATGGCACTCATAAAGGGAGTGTGG----CTCAGAAAG 99272  
Oy 2917 GAGGAAGAGGAAAGGCCCTTGTGCCCACTGAGCGAGGCTCATCTGATGAGAGAT 2976  
Db 99273 GAGGAGGACTGGAAAGGTCTTGTGCCCACTGAGCGAGGCTCATCTGATGAGAGAT 99332  
Oy 2977 CTGCGGGGTGGCAGAGCCCACTGCTGTCTGCCAAGGAAACCCCAAGTGAAGTCTG 3036  
Db 99333 CTGAAGAGGTGCCAGGGGCCAC--CTGTCCCAAGGAAACCCCAAGTGAAGTCTG 99388  
Oy 3037 GCTTGGGTGCTGATTCAGCTTACAGACCCAGAGTCTACTCATCCCATCCAGT 3096  
Db 99389 GCCTTGGGTGCTGATTCAGCTTACAGTGCAGAGGCCCAAGAGTCTGTCCCATCCATCTCGTG 99448  
Oy 3097 GC----- 3098  
Db 99449 CCCCANN 99508  
Oy 3099 ----- 3098  
Db 99509 NNN 99568  
Oy 3099 CCCCCTGGCCCGCACACCCCAACCCCGACCTCCCTGCACTCTCTAGGGCTGAGG-- 3156  
Db 99569 CCCCCACCCACACCATCCCTCCCCCAACCCCGTACTTCTTCTAGGGAGGAGGT 99628  
Oy 3157 -----GTGGCCAG 3164  
Db 99629 GCTCTGTAGTTTCCCATCTTGGCGGGAAAGGAAAGATGTTCCGGGACCTGTGGCCTG 99688  
Oy 3155 CCCTGTGGG--GGTTGCCCTACCTCAGGTAGAGCCCAAGTCTTACCGGAAAGTGCACCC 3222  
Db 99689 CCCTGTGGGAAAGATGCTTACTCTCAGGAGAGCTCAGGTCTTACGGGGAAGTGCACCC 99748  
Oy 3223 CATCCCTGAAGCTGAGAGCCAAAGGGGGGAGCACAGCGAG--CTCAGGCTGTGAGGC 3278  
Db 99749 CATCCCTGAAGCTGAGAGCCAGAGGCGGAGCAGCACGCGCAGCTNTGCTCAAGNC 99808  
Oy 3279 TGTGTGTGGCTCTAGTTTCCAGGAGCTTGGGACCTT--CCCCACCCCCCATCT 3336  
Db 99809 TGTGTGTGGCTCTAGTTTCCAGGAGCTTGGGACCTT--CCCCACCCCCCATCT 99868  
Oy 3337 ATCTCTCTGGGGCCCTATCTTCCCTTATATG--TGAAGGAAGTTCTCTGGGGGGGGG 3394  
Db 99869 ATCTCTCTGGGGCCCTATCTTCCCTTATATGNGTGAAGGAAGTTCTCTGGGGGGGGG 99928  
Oy 3395 GTGGGTGAGAGCAAAAGTCTGTGGTCTCTGAGCAGCTTCCCAACTTCTTAAG 3454  
Db 99929 -----GTGAGAGCAAAAG-----TGGTCTCTCTGAGCAGCTTCTCCCAACTTCTTAAG 99979  
Oy 3455 ATCTCCAGAGTGTGGCTGTCTTCCAGACAGTAAAGCAATTGGTGGGAGACATGG 3514  
Db 99980 ATCTCCAGAGTGTGGCTGTCTTCCAGACAGTAAAGCAATTGGTGGGAGACATGG 100039  
Oy 3515 TGACACAGAGTGTGGAGGGAGAGGCTCTTCTCTCTCTGAGCAGCTGTCTTCT 3574  
Db 100040 TGACACAGAGTGTGGAGGGAGAGGCTCTTCTCTTCTTCTTCTGCACTGTGAC----- 100094

Oy 3575 GTAGCACCTGTGATTAAGTTTGGGGGTGAGTAAGTGTCTGAAACTCTGAAGAAGCA 3634  
Db 10095 -----AGTTAGTGAAGGGGGGTGAGTAAGTATTCGGAACTCTGAAGAAGCA 100145  
Oy 3635 AGAAGCAGCAGGCTGTCTTGGGCTTCAATGAAGAAAGTTG--ACAGACCCCTTCTCT 3692  
Db 100146 AGAAGCTAGACACTGTCTTGGGCTTCAATGAAGAAAGTTGCTTCCCTGTAAACACCT 100205  
Oy 3693 GTAACTCAGCTTCTCATCTGTAGATTTCCCTGGGAGCCAAAGTGGCTCTGGGACTC 3752  
Db 100206 TCACCTCAGGTTGGGTCCATCTGTGACAGATTCCTGGGACTGA--GTGGCTCTGGGACTC 100264  
Oy 3753 AGATTTCACATTTAAAA--TCAGACAGTCTCTGAG-----ACTTGGCTCG 3798  
Db 100265 AGATTTCACATTTAAAAATCTGAGACAGTCTCTGAGAAAGCTGAATTAAGTGTGCTCT 100324  
Oy 3799 TGCTGTATTACTTCTCTCTGCTGCTGCTCATTTCTGTCTTCAATGTCATCATCTG 3858  
Db 100325 TGCGAACAAATTTCTTCTCTCTCTGAGGTTCAATTTCTTGTATGTCCTCACCTTCTG 100384  
Oy 3859 AAATGG-----TTTCTTGTGTACCATTCCTCCGACACTCTGGGAGGTCATCTTGG 3914  
Db 100385 AAATGGTTTCTTGTGTGTACCATTCCTCCGACACTCTGAGAGGTTGGCTCTTGG 100444  
Oy 3915 CAC-----ATGTATCCCTGGATGTAACTGACCCACAGAG--GAGAGGGGAGAGTCAAG 3969  
Db 100445 CCCACGATGATCTCTGGATGTAACTGACATCACACAGAGGGAGAAAGCAAGTCTG 100504  
Oy 3970 AG-----CTGTGTCTAGAGCCCTATTAGGCTGAGCATCACCCCTTCTTCTAGAAATGCC 4025  
Db 100505 AGGCTACAGGCTCTAGAGCCCTCTGCTGGCTGTGATCTCTGCTCCCTTAAGAAAGGCC 100564  
Oy 4026 CTTCATTTTGGGTATACATGATATTTTATATCAGAGTGGGAGTGAACCAACCT 4085  
Db 100565 CTCTCTTCTCAGTCTCGGTGTCTATTT--TATAGACTACAGTGAACCAACCT 100622  
Oy 4086 GCCCAAGATTGGGACTACATCAGACCAAGGTTATCTGCTCAGAAATCCCTGTCACT 4145  
Db 100623 GCCCAAGATTGGGACTACATCAGACCAAGGTTATCTGCTCAGAAATCCCTGTCACT 100682  
Oy 4146 T--GAGTTGGGAATCTGCTCTGGGGCTTCAAGGCTTGTGTTAGCAGAGGATCTC 4204  
Db 100683 TGGAGTTGGGAGTATCTCCCTCAAGGGCTCCAGGGCTGTGTTAGCAGAGGCAATCTC 100742  
Oy 4205 TTTGTATAG--GCATGACCTATGCTATGAGTGTACTATCTTCTCTCAATTTAAAGC 4261  
Db 100743 TTTGTATAGGAGCTGTACCTAACCCATGAGTGTACTACACTCTCTCTCAATTTAAAGC 100802  
Oy 4262 TGGACTTAAACCCACAGCGAGCGCCAGATTTCTTACAGTTGTACCCCAAGCAACAA 4321  
Db 100803 TGGACTTAAACCCCTGGAGCAGCAGCAGATTTCTCTTAATGTGTGACTCCAAAGCAACG 100862  
Oy 4322 GACAGTATATGACAGGATAGTACTGGGGAGAAAGAACTTAA--ACCCCCCAAG 4380  
Db 100863 GACAGTATATGAGGAGCTGACCTAACCCATGAGTGTACTACACTCTCTCTCAATTTAAAGC 100922  
Oy 4381 GCCCAGAGTTCCTGCTTCTTCAATGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 4432  
Db 100923 AGGTCT 100982  
Oy 4433 ATGGGCTGTGAGTGTGAGCTTCAAGCATGATGA--TGTCTATGTGTAGTGTATTA 4491  
Db 100983 ATGGGCTGTGAGTGTGAGCTTCAAGCATGATGA--TGTCTATGTGTAGTGTATTA 101042  
Oy 4492 TCTGAG--CACTTGGAGGCTGAAGCAGAGATCTTATGTATTGAGGCCAGCTGAGC 4550  
Db 101043 TCTGAGCACTTGAAGGCGCCGAAGCATG-----TTTGTATGTATTGAGGCCAAGCTGAGC 101097  
Oy 4551 TATAGAGGAGACTTGTCTTTAAGAAAAAATGAAGCCCAAGCAGTGTGTGACACAGCC 4610  
Db 101098 TATAGAGGAGAGCTGTCTTTAAGAAAAAAGAGCAACCAACT----- 101144







AC119514	LOCUS	AC119514	203605 bp	DNA	linear	HTG 19-NOV-2002		
AC119514	DEFINITION	Rattus norvegicus clone CH230-407A8, ***	SEQUENCING IN PROGRESS					
AC119514	ACCESSION	AC119514						
AC119514	VERSION	AC119514.5	GI:25090934					
AC119514	KEYWORDS	HTG: HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.						
AC119514	SOURCE	Rattus norvegicus (Norway rat)						
AC119514	ORANISM	Rattus norvegicus						
AC119514	ORANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.						
REFERENCE	AUTHORS	1 (bases 1 to 203605) Muzny,D.,Marie., Metzker,M.,Lee., Abramzon,S., Adams,C., Alder,J., Allen,C., Allen,H., Alsbrooks,S., Antin,A., Anguiano,D., Anyalebechi,Y., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H., Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Bernabed,F., Biswalto,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M., Bryant,N., Buhay,C., Burck,P., Burrell,K., Calderon,E., Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J., Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L., Davila,M.L., Davis,C., Davy-carroll,L., De Andra,C., Dedetich,D., Delgado,O., Denison,S., Detramo,C., Ding,Y., Dinh,H., Diya,K., Draper,H., Dugan-Rocha,S., Dunn,A., Dublin,K., Duvall,B., Eaves,K., Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G., Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P., Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M., Georgeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guenara,W., Gunaratne,P., Haaland,M., Hamil,C., Hamilton,C., Hamilton,K., Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J., Hernandez,B., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M., Hollins,B., Howells,S., Hulik,S., Hume,J., Idlebird,D., Jackson,A., Jackson,L., Jacob,L., Jiang,H., Johnson,Z., Johnson,R., Jolivet,A., Karpachy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C., Kwis,C., Krift,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J., Liu,J., Liu,Y., Liu,Y., London,P., Longacre,S., Lopez,J., Lorenzenhwa,L., Louisedge,H., Lozado,R.J., Lu,X., Ma,J., Mahalingam,B., Mahindaratne,M., Mahmood,M., Malloy,K., Mangun,A., Mangun,B., Mapua,P., Martin,K., Martin,R., Martinez,E., Mahoney,S., McLeod,M.P., McNeill,T.Z., Meenen,E., Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S., Morgan,M., Morris,K., Morris,S., Mundasa,M., Murphy,M., Natr,L., Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Nwokolienh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K., Patenack,S., Paul,H., Perez,A., Perez,L., Pflankoch,C., Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L., L., Puzo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R., Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F., Rives,C., Rodey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J., Sanders,W., Savery,G., Scherer,S., Scott,G., Shatman,S., Shen,H., Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D., Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J., Steidle,M., Strong,R., Sutton,A., Svatlek,A., Tabac,P., Taylor,C., Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K., Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J., Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F., Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K., Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V., Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O., Weinstock,G. and Gbbs,R.A.						
JOURNAL	TITLE	Unpublished						
REFERENCE	AUTHORS	2 (bases 1 to 203605) Worley,K.C.						
JOURNAL	TITLE	Direct Submission						
REFERENCE	AUTHORS	Submitted (28-APR-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA						
REFERENCE	AUTHORS	3 (bases 1 to 203605) Rat Genome Sequencing Consortium.						
JOURNAL	TITLE	Direct Submission						

JOURNAL

Submitted (19-NOV-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA

On Nov 19, 2002, this sequence version replaced gi:23913694.  
The sequence in this assembly is a combination of BAC based reads  
and whole genome shotgun sequencing reads assembled using Atlas  
(<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described  
in the feature table below represents a scaffold in the Atlas  
assembly (a 'contig-scaffold'). Within each contig-scaffold,  
individual sequence contigs are ordered and oriented, and separated  
by sized gaps filled with Ns to the estimated size. The sequence  
may extend beyond the ends of the clone and there may be sequence  
contigs within a contig-scaffold that consist entirely of whole  
genome shotgun sequence reads. Both end sequences and whole genome  
shotgun sequence only contigs will be indicated in the feature  
table.

----- Genome Center

Center : Baylor College of Medicine  
Center code : BCM  
Web site : <http://www.hgsc.bcm.tmc.edu/>  
Contact : hgsc-help@bcm.tmc.edu

----- Project Information

Center Project name : GOUV  
Center Clone name : CH230-407A8

----- Summary Statistics

Assembly program : Phrap; Version 0.990329  
Consensus quality : 164885 bases at least Q40  
Consensus quality : 169466 bases at least Q30  
Consensus quality : 172574 bases at least Q20  
Estimated insert size : 164510; sum-of-contigs estimation  
Quality coverage : 6x in Q20 bases; sum-of-contigs estimation

-----

\* NOTE: Estimated insert size may differ from sequence length  
( see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)) .  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 8 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

    1          97404: contig of 97404 bp in length

        \*      97405: gap of unknown length

        \*      97505: 108574: contig of 11070 bp in length

        \*      108575: 108675: gap of unknown length

        \*      112656: 112655: contig of 3981 bp in length

        \*      112756: 112755: gap of unknown length

        \*      150903: 150902: contig of 38047 bp in length

        \*      150903: 150902: gap of unknown length

        \*      200023: 200022: gap of 49121 bp in length

        \*      200124: 200123: gap of unknown length

        \*      201124: 201202: contig of 1079 bp in length

        \*      201203: 201302: gap of unknown length

        \*      201303: 202347: contig of 1045 bp in length

        \*      202348: 202447: gap of unknown length

        \*      202448: 203605: contig of 1158 bp in length.

FEATURES

source

    1..203605

        /organism="Rattus norvegicus"  
        /mol\_type="genomic DNA"  
        /db\_xref="taxon:10116"  
        /clone="CH230-407A8"

    1..2247

        /note="wgs\_end\_extension  
        clone\_end:Sp6"  
        1773..2248  
        /note="clone\_boundary  
        clone\_end:Sp6  
        site:  
        end\_sequence:BZ153243"  
        complement(33359..34344)  
        /note="clone\_boundary

misc\_feature

misc\_feature

misc\_feature





Db 1759 GTGATGTGTGTCGATTTACACACACACTATATTATTTTAACTCAGAGCCTTTCTGT 181

Steinle, M., Strong, R., Sutton, A., Svatek, A., Taber, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K.,

	misc-feature	121744..143417	/note=Clone_Boundary clone_end:Sp6 site:EcoRI end_sequence:BH262099"
BASE COUNT	55073 a	53210 c	54594 g 58837 t 18721 others
ORIGIN			
Query Match	2.1%;	Score 188.6;	DB 2; Length 240435;
Best Local Similarity	62.4%;	Pred. No. 5.5e-43;	
Matches 384;	Conservative 0;	Mismatches 214;	Indels 17; Gaps 5;
OY	1799	ATAATAGTTCACAAATGATATATATATATAGTAATCTTTTAAAAAGTATATGTGGAGG	1858
Db	83343	ATTATTATBTACCACACAAGAACAATATATATCACTTTGCCAACAATGTTGGAGACTG	8328
OY	1859	GAGAGATGGCTCAGCTTCCAGAGACACACTGCTGCTTCGCAGAGAGACCTAGATTCACTTC	1918
Db	83283	GAGAGACACTATACAGCTTAGACAGACAGCGCCGCTTGCCAGAGAGCCCTGTTCACTTC	83222
OY	1919	CCAGGACCTCATATGTGGCTCACAGCCATCTGTAAATCCAGTTTCCAAAGGTTCCACACC	1978
Db	83223	CCAGAACCCACACGGTGTACTACACACCATTTGTATCTGTATGCCAGGGGCCCAACACC	8316
OY	1979	CTCTTCTGGCCTCCACAGGACCCACATATACATAGTACACAGACATACATGACGCAAAACA	2038
Db	83163	CTCTTCTGGCATCTGTGTGTGCACACAGACATGACACAGGATATACAGACATCCAGG-AAACAA	8310
OY	2039	CCCATACACACATATATATATAGGAACCTTAAAAGTGCACTGTGTTGTTAAACATTTGTG	2098
Db	83104	CTCACACATPAGAACTGTAGATATCATTTTTTTTGTATGTGTGTTCTACMACTATGGA	8304
OY	2099	CTTACACATGCTGATTTGGAACA---TGTACACAGCACACACT---GAGAGGAGATGTGG	2152
Db	83044	TGCAGCCACGAGAACCATGACATTTACTGGGGGTGTGAGGGACAGACTTAAAAAGAAATTTTG	8298
OY	2153	GGCTGTGAGAGATGGCTCAGCGTTTAAAGACACTGACTCTTTCGGAAGAGAGTCTCGA	2212
Db	82984	GGCTGTGAGAAATGGCTCAGATTTAAATAGCACTGAGTGTCTCC-----AGAGGTCTGGA	8292
OY	2213	GTTCAAATCCGAGAACACATGGTGGCTCAACAACATCATATATGATATGATGACACCT	2272
Db	82928	GTTCAAATCCGAGAACACATGGTGGCTCAACAACATCTGTATATGATGTGATGTCCT	8286
OY	2273	CTTCTGTGTCATCTGAGAGACAGCTGCAGAGCTACAGTGTACTTATGATATCTATATAATA	2332
Db	82868	CTTCTGTGTCATCTGAGAGACAGCTGCAGAGCTACAGTGTACTTATGATATCTATATAATA	8281
OY	2333	AATCTTTTAAAAAATGAAGGGGATCTGAGACACCTCAAAGAGATTATGAGCAGT	2392
Db	82814	ATCTTTTAAAAAAGAGAAATATATATGCTTCTCTACATGTGGACTTTAGGTTT	8275
OY	2393	GACTCAGCGGTGAT 2407	
Db	82754	TTTTTAAGATTTATT 82740	
RESULT 10			
ACI32539/c	ACI32539	262322 bp	DNA linear HTG 19-NOV-2002
LOCUS	Rattus norvegicus clone CH230-196N8,	WORKING DRAFT SEQUENCE, 3	
DEFINITION	unordered pieces.		
ACCESSION	ACI32539		
VERSION	ACI32539.3	GI:25072816	
KEYWORDS	HTG; HTGS-PHASE1; HTGS-DRAFT; HTGS-FULLTOP.		
SOURCE	Rattus norvegicus (Norway rat)		
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus;		
REFERENCE	1 (bases 1 to 262322)		
AUTHORS	Mundy,D.,Marle,, Metzker,M.,Lee,, Abramzon,S., Adams,C., Alder,J., Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,		



Ayvazbechi V., Aoyagi A., Ayodeji M., Baca E., Baden H.,  
 Baldwin D., Bandaranaike D., Barber M., Barnstead M., Bernhied F.,  
 Biswalio K., Blair J., Blankenburg K., Blyth P., Brown M.,  
 Bryant N., Buhay C., Burch P., Butrell R., Calderon E.,  
 Cardenas V., Carter K., Cavazos I., Ceasar H., Center A.,  
 Chacko J., Chavez D., Chen G., Chen R., Chen Y., Chen Z., Chu J.,  
 Cleveland C., Cockrell R., Cox C., Coyle M., Cree A., d'Souza D.,  
 Davila M.L., Davis C., Davy-Carroll J., De Anda C., Dederich D.,  
 Delgado O., Denison S., Detramo C., Ding Y., Dinh H., Divya K.,  
 Diaper H., Dugan-Rocha S., Dunn A., Durkin K., Duvall B., Eaves K.,  
 Egan A., Escotto M., Eugene C., Evans C.A., Falls T., Fan G.,  
 Fernandez S., Finley M., Flagg N., Forbes L., Foster M., Foster P.,  
 Fraser C.M., Gabisi A., Ganta R., Garcia A., Garner T., Garza M.,  
 Gebregeorgis E., Geer K., Gill R., Gady M., Guerra W., Guevara W.,  
 Gunaratne P., Haaland W., Hamill C., Hamilton C., Hamilton K.,  
 Harvey Y., Havlak P., Hawes A., Henderson N., Hernandez T.,  
 Hernandez S., Hines S., Hladun S.L., Hodgson A., Hogues M.,  
 Hollins B., Howells S., Hulik S., Hume J., Idlebird D., Jackson A.,  
 Jackson L., Jacob L., Jiang H., Johnson B., Johnson R., Jolyvel A.,  
 Kapachy S., Kelly S., Kelly S., Khan Z., King L., Kovar C.,  
 Kowis C., Kraft C.L., Lebow H., Levan J., Lewis L., Li Z., Liu J.,  
 Liu J., Liu M., Liu Y., London P., Longacre S., Lopez J.,  
 Lorenzshewa L., Louiseged H., Lozardo R.J., Lu X., Ma J.,  
 Maheshwari M., Mahndartine M., Mahmood M., Malloy K., Mangum A.,  
 Mangum B., Mapua P., Martin K., Martin R., Martinez E.,  
 Manthey S., McLeod M.P., McNeill T.Z., Meenen E.,  
 Milosavljevic A., Miner G., Minja E., Montemayor J., Moore S.,  
 Morgan M., Morris K., Morris S., Muntadasa M., Murphy M., Natl L.,  
 Narkavici S., Neal D., Newton N., Nguyen N., Norris S.,  
 Nweoklemeh O., Okunou G., Olampunsgoon A., Pal S., Parks K.,  
 Pasternak S., Paul H., Perez A., Perez L., Pfankoch C.,  
 Plopper F., Polndexter A., Popovic D., Primus E., Pu L., L.,  
 Plazo M., Quiroz J., Rachlin E., Reeves K., Reiger M.A., Reigh R.,  
 Reilly B., Reilly M., Ren Y., Reuter M., Richards S., Riggs F.,  
 Rives C., Rodkey R., Rojas A., Rose M., Rose R., Ruiz S.J.,  
 Sanders W., Saverly G., Scherer S., Scott G., Shatsman S., Shen H.,  
 Shetty J., Shvartsbeyn A., Sisson I., Sitter C.D., Smajz D.,  
 Sneed A., Sodergren E., Song X.-Z., Sorelle R., Sosa J.,  
 Steidle M., Strong R., Sutton A., Syatel K.A., Taber P., Taylor C.,  
 Taylor T., Thomas N., Thomas S., Tingey A., Trejos Z., Umanu K.,  
 Valas R., Vera V., Villasana D., Waldron L., Walker B., Wang J.,  
 Wang O., Wang S., Warren J., Warren R., Wei X., White F.,  
 Williams G., Willson R., Wleczky R., Wooden H., Wotley K.,  
 Wright D., Wright R., Wu J., Yakub S., Yen J., Yoon L., Yoon V.,  
 Yu F., Zhang J., Zhou J., Zhou X., Zhao S., Dunn D., von  
 Niederhausen A., Weiss R., Smith D.R., Holt R.A., Smith H.O.,  
 Weinstein G. and Gibbs R.A.  
 Direct Submission  
 Title  
 Journal  
 Unpublished  
 2 (bases 1 to 262322)  
 Rat Genome Sequencing Consortium.  
 Direct Submission  
 Submitted (01-SEP-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 3 (bases 1 to 262322)  
 Rat Genome Sequencing Consortium.  
 Direct Submission  
 Submitted (19-NOV-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 COMMENT  
 The sequence in this assembly is a combination of BAC based reads  
 and whole genome shotgun sequencing reads assembled using Atlas  
 (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described  
 in the feature table below represents a scaffold in the Atlas  
 assembly (a 'contig-scaffold'). Within each contig-scaffold,  
 individual sequence contigs are ordered and oriented, and separated  
 by sized gaps filled with Ns to the estimated size. The sequence  
 may extend beyond the ends of the clone and there may be sequence  
 contigs within a contig-scaffold that consist entirely of whole  
 genome shotgun sequence reads. Both end sequences and whole genome  
 shotgun sequence only contigs will be indicated in the feature

	table.
	----- Genome Center
	Center: Baylor College of Medicine
	Center code: BCM
	Web site: http://www.hgsc.bcm.tmc.edu/
	Contact: hgsc-help@bcm.tmc.edu
	----- Project Information
	Center project name: GS0X
	Center clone name: CH230-196N8
	----- Summary Statistics
	Assembly program: Phrap; version 0.990329
	Consensus quality: 215928 bases at least Q40
	Consensus quality: 218585 bases at least Q20
	Consensus quality: 220210 bases at least Q20
	Estimated insert size: 221226; sum-of-contigs estimation
	quality coverage: 7x in Q20 bases; sum-of-contigs estimation
	-----
	* NOTE: Estimated insert size may differ from sequence length
	* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_drift_data.html).
	* NOTE: This is a 'working draft' sequence. It currently
	* consists of 3 contigs. The true order of the pieces
	* is not known and their order in this sequence record is
	* arbitrary. Gaps between the contigs are represented as
	* runs of N, but the exact sizes of the gaps are unknown.
	* This record will be updated with the finished sequence
	* as soon as it is available and the accession number will
	* be preserved.
	* 1 53758: contig of 53758 bp in length
	* * 53759 53858: gap of unknown length
	* * 53859 253122: contig of 199264 bp in length
	* * 253123 253222: gap of unknown length
	* * 253223 262322: contig of 9100 bp in length.
	* Location/Qualifiers
	1. 262322
FEATURES	/organism="Rattus norvegicus"
SOURCE	/mol_type="genomic DNA"
	/db_xref="taxon:10116"
	/clone="CH230-196N8"
	3152..4104
misc_feature	/note="clone boundary
	clone_end:Sp6
	site:
	end_sequence:BH320154"
	53859..54886
misc_feature	/note="wgs_contig"
	91022..92334
misc_feature	/note="wgs_contig"
	complement(193497..194191)
	/note="clone boundary
	clone_end:T7
	site:
	end_sequence:BH320153"
misc_feature	251042..253122
	/note="wgs_end_extension
BASE COUNT	56211 a 52467 c 54612 g 58218 t 40814 others
ORIGIN	
Query Match	2.1%; Score 188.4; DB 2; Length 262322;
Best Local Similarity	58.4%; Pzed.No.6.4e-43;
Matches 395; Conservative	0; Mismatches 261; Indels 20; Gaps 6;
QY	1799 ATATAGTTACATGATTAATAATTAGTAAATCTTTAAAAGGTATATGTTGGAGG 1858
Dd	178963 ATTATTTATGAGCCACAACAATAATGATTCATTTGCCAACAAATGTTGGGAAGCTC 178904
QY	1859 GAGAGTGCGTCAGCTTCAGAGCAGCATTTGCTCTTGCAAGAGAGCCTAGATTCAGTTG 1918
Dd	178903 GAGAGCGACTAAGCAGCTTAAGAAGCAGTGGCCGCTCTGCAGAGAGACCCTGTTCAAGTTC 178844
QY	1919 CCAGGAGCTATATGTTGGCTCACAGCCATCTGTAATCCATTTCCAGAGGTTCCACAGC 1978
Dd	178843 CACGACACCCACAGCGGACTCTCAACCAATTTGTACTTAGTCCAGGAGGCCCAAGAC 178784

DB	Accession	Version	KeyWords	Source	Organism	Reference	Authors
OY	1979	1	CTCTCTTGCGCCCTCCACAGGCAACCATACATACATAGTACACAGACTATACATGACGCAAAACA				
DB	178783	1	CTCTTCTGGCACTGTGTGTCACACGACATGACACACGGTATACAGGATCTCAGG-AAAACA				
OY	2039	1	CCCATATACACACATTAATAATTAAGGAAATCTTAAAGAGTATGTTGGTAACATTTGTG				
DB	178724	1	CTCACACATACGCACTGTAGTAAATCATTTTTTTAGTTATGTTTGGTACAAAGTATGCA				
OY	2099	1	CTTACACATGCTGATTTGAAGACA---TGTACAACGCACACACT--GAAGAGGATCTGG				
DB	178664	1	TGCACCCAGAGACCATGACATTACTGGGGGTAGGGGACAGACTCAAAAAGAGAAATATTG				
OY	2153	1	GCGCTGAGAGATGGCTGACAGCGGTTTAAAGACACTGATGCTGCTCCGAGAAAGAGTCTG				
DB	178604	1	GCGTGGAGAAATGGCTCACTAGTATTATTAATACACTGACTGCTCTCCAG---AGGTCTGA				
OY	2213	1	GTTCAAAATCTTAGCAACCATGATGGCTCACAAACCATCATTAATAGATCTTGACACCT				
DB	178548	1	GTTCAAAATCTTAGCAACCATGATGGCTCACAAACCATCATTAATAGATCTTGACACCT				
OY	2273	1	CTTCGGTGACATCTGAAGACAGCTGACAGCTACAGCTGATGATTAATTAATTAATA				
DB	178488	1	CTTCGGTGATGCTGTAAGAC---AGTACATGATGATCTATACATTAATAATA				
OY	2333	1	AATCTTTTAAAAAATGAAGAGGATCTAGACACTGACACTCAAAAGAGATTAATGACACT				
DB	178437	1	AAAATCTTTTAAAAAAGAGAAATATTATATGCTNNNNNNNNNNNNNNNNNNNNNNNN				
OY	2333	1	GACTCAGCGGATTAATATATCTGATGAGTTTTCCTTCGCTGGCTTGCAACTGGGCTG				
DB	178377	1	NN				
OY	2453	1	GACACAGCGCCCTTTT 2468				
DB	178317	1	TATGACTCTCTATCTT 178302				
RESULT 11	AC114361	238330 bp	DNA	linear	HTG 22-SEP-2002		
LOCUS	AC114361	238330 bp	DNA	linear	HTG 22-SEP-2002		
DEFINITION	Rattus norvegicus clone CH230-45C9, *** SEQUENCING IN PROGRESS ***						
ACCESSION	AC114361						
VERSION	AC114361.4	GI:23195300					
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.						
SOURCE	Rattus norvegicus (Norway rat)						
ORGANISM	Rattus norvegicus						
	Eumetazoa; Chordata; Craniata; Vertebrata; Euteleostomi;						
	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;						
	Rattus.						
REFERENCE	1 (bases 1 to 238330)						
AUTHORS	Munry,D,Merle,,Metzker,M,Lee,,Abramson,S,,Adams,C,,Alder,J.,						
	Allen,C,,Allen,H,,Alsbrooks,S,,Amlin,A,,Anguiano,D.,						
	Anyalebech,V.,Aoyagi,A.,Ayodeji,M.,Baca,E.,Baden,H.,						
	Baldwin,D.,Bandarainaike,D.,Barber,M.,Barnstead,M.,Behnmed,F.,						
	Biswalto,K.,Blair,J.,Blankenburt,K.,Blyth,P.,Brown,M.,						
	Bryant,N.,Buhay,C.,Burich,P.,Buttelli,K.,Calderon,E.,						
	Cardenas,P.,Carter,K.,Cavazos,I.,Ceasar,H.,Center,A.,						
	Chacko,J.,Chavez,D.,Chen,G.,Chen,R.,Chen,Y.,Chen,Z.,Chu,J.,						
	Cleveland,C.,Cockrell,R.,Cox,C.,Coyle,M.,Cree,A.,D'Souza,L.,						
	Devila,M.L.,Davis,C.,Davy-Carroll,L.,De Anda,C.,Dedrich,D.,						
	Delgado,O.,Denson,S.,Deramo,C.,Ding,Y.,Dinh,H.,Ditya,K.,						
	Diaper,H.,Dugan-Rocha,S.,Dunn,A.,Durbin,K.,Duval,B.,Eaves,K.,						
	Egan,A.,Escotto,M.,Eugene,C.,Evans,C.A.,Falls,T.,Fan,G.,						
	Fernandez,S.,Finley,M.,Flaggs,N.,Forbes,L.,Foster,M.,Foster,P.,						
	Fraser,C.M.,Gabisi,A.,Gatta,R.,Garcia,A.,Garner,T.,Garza,M.,						
	Gebegeorgis,E.,Geer,K.,Gilli,R.,Grady,M.,Guerra,W.,Guevara,W.,						
	Gunaratne,P.,Haaland,W.,Hamil,C.,Hamilton,C.,Hamilton,K.,						
	Harvey,Y.,Hayak,P.,Hawes,A.,Henderson,N.,Hernandez,J.,						
	Hernandez,R.,Hines,S.,Hladun,S.L.,Hodgson,A.,Hogues,J.,						
	Hollins,B.,Howells,S.,Hulyk,S.,Hune,J.,Idlebird,D.,Jackson,A.,						
	Jackson,L.,Jacob,L.,Jiang,H.,Johnson,B.,Johnson,R.,Jolivet,A.,						

KOVAC, C., KRAFT, C.L., LEDOW, H., LEVAN, J., LEWIS, L., LI, Z., LIU, J.,  
 LIU, J., LIU, W., LIU, Y., LONDON, P., LONGACRE, S., LOPEZ, J.,  
 LORENSUNHEVA, L., LOUISEGED, H., LOZADO, R.J., LU, X., MA, J.,  
 MAHESHWARI, M., MAHINDARTINE, M., MAHMUD, M., MALLORY, K., MANGUM, A.,  
 MANGUM, B., MAPUA, P., MARTIN, K., MARTIN, R., MARTINEZ, E.,  
 MATHNEY, S., MCLEOD, M.P., MCNEILL, T.Z., MEENE, E.,  
 MILOSEVJELIC, A., MINER, G., MINJA, E., MONTENAYOR, J., MOORE, S.,  
 MORGAN, M., MORRIS, K., MORRIS, S., MUNDASA, M., MURPHY, M., NAIR, L.,  
 NAKKEVIS, C., NEAL, D., NEWTON, N., NGUYEN, N., NORRIS, S.,  
 NAWOKLEMEH, O., OKWUONU, G., OIARUNPUSAGGON, A., PAL, S., PARKS, K.,  
 PATERNAK, S., PAUL, H., PEREZ, A., PEREZ, L., PFANKOCH, C.,  
 PLOPPER, F., POINDEXTER, A., POPOVIC, D., PRINS, E., PU, L.-L.,  
 PUZZO, M., QUIROZ, J., RACHLIN, E., REEVES, K., REGIER, M.A., REIGH, R.,  
 RELLY, B., REILLY, M., REN, Y., REUTER, M., RICHARDS, S., RIGGS, F.,  
 RIVES, C., RODEY, T., ROJAS, A., ROSE, M., ROSE, R., RUIZ, S.J.,  
 SANDERS, W., SAVERY, G., SCHERER, S., SCOTT, G., SHATMAN, S., SHEN, H.,  
 SHETTY, J., SHVARTSBEYN, A., SISON, I., SITTER, C.D., SMAJS, D.,  
 SNEED, A., SODERGREN, E., SONG, X.-Z., SORELLE, R., SOSA, J.,  
 STEIMLE, M., STRONG, R., SUTTON, A., STAYLEK, A., TABOR, P., TAYLOR, C.,  
 TAYLOR, T., THOMAS, N., THOMAS, S., TINGEY, A., TREJOS, Z., USMANI, K.,  
 VALAS, R., VERA, V., VILLASANA, D., WALDRON, L., WALKER, B., WANG, J.,  
 WANG, Q., WANG, S., WARREN, J., WARREN, R., WEI, X., WHITE, F.,  
 WILLIAMS, G., WILSON, R., WLECZYK, R., WOODEN, H., WOOLEY, K.,  
 WRIGHT, D., WRIGHT, R., WU, J., YAKUB, S., YEN, J., YOON, L., YOON, V.,  
 YU, F., ZHANG, J., ZHOU, J., ZHOU, X., ZHAO, S., ZUNN, D., von  
 NIEDERHAUSEN, A., WEISS, R., SMITH, D.R., HOLT, R.A., SMITH, H.O.,  
 WEINSTECK, G. and GIBBS, R.A.  
 Direct Submission  
 Unpublished  
 2 (bases 1 to 238330)  
 Worley, K.C.  
 Direct Submission  
 Submitted (08-MAR-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 3 (bases 1 to 238330)  
 Rat Genome Sequencing Consortium.  
 Direct Submission  
 Submitted (22-SEP-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 On Sep 19, 2002 this sequence version replaced gi:21744367.  
 The sequence in this assembly is a combination of BAC based reads  
 and whole genome shotgun sequencing reads assembled using Atlas  
 (<http://www.hgsc.bcm.tmc.edu/projects/raft/>). As a result, the  
 sequence may extend beyond the ends of the clone and there may be  
 contigs that consist entirely of whole genome shotgun sequence  
 reads. Both end sequences and whole genome shotgun sequence only  
 contigs will be indicated in the feature table.  
 ----- Genome Center  
 Center: Baylor College of Medicine  
 Center code: BCM  
 Web site: <http://www.hgsc.bcm.tmc.edu/>  
 Contact: hgsc-help@bcm.tmc.edu  
 ----- Project Information  
 Center project name: GRHV  
 Center clone name: CH230-45C9  
 ----- Summary Statistics  
 Assembly program: Phrap; version 0.990329  
 Consensus quality: 218627 bases at least Q40  
 Consensus quality: 221500 bases at least Q30  
 Consensus quality: 223329 bases at least Q20  
 Estimated insert size: 234329; sum-of-contigs estimation  
 Quality coverage: 5x in Q20 bases; sum-of-contigs estimation  
 -----  
 \* NOTE: Estimated insert size may differ from sequence length  
 \* (see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_diff\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_diff_data.html))  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 3 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.



*	11436	14325:	contig of 2890 bp	in length
*	14326	14425:	gap of 100 bp	
*	14428	17305:	contig of 2880 bp	in length
*	17306	17405:	gap of 100 bp	
*	17406	20166:	contig of 2761 bp	in length
*	20167	20266:	gap of 100 bp	
*	20267	23474:	contig of 3308 bp	in length
*	23475	23574:	gap of 100 bp	
*	23575	26403:	contig of 2829 bp	in length
*	26404	26503:	gap of 100 bp	
*	26504	28899:	contig of 3396 bp	in length
*	28900	29999:	gap of 100 bp	
*	30000	34929:	contig of 4930 bp	in length
*	34930	35029:	gap of 100 bp	
*	35030	41463:	contig of 6634 bp	in length
*	41464	41563:	gap of 100 bp	
*	41564	47890:	contig of 6327 bp	in length
*	47891	47990:	gap of 100 bp	
*	47991	56092:	contig of 8102 bp	in length
*	56093	56192:	gap of 100 bp	
*	56193	63063:	contig of 6671 bp	in length
*	63064	63163:	gap of 100 bp	
*	63164	70819:	contig of 7656 bp	in length
*	70820	70919:	gap of 100 bp	
*	70920	82482:	contig of 11563 bp	in length
*	82483	82582:	gap of 100 bp	
*	82583	98645:	contig of 16063 bp	in length
*	98646	98745:	gap of 100 bp	
*	98746	117782:	contig of 19037 bp	in length
*	117783	117882:	gap of 100 bp	
*	117883	149425:	contig of 31543 bp	in length

```
misc_feature      /clone_11b=RPCT-11 Human Male BAC
                  1..8144
                  /note="assembly-fragment
clone_end:SP6
vector_side:left"
misc_feature      8245..9350
                  /note="assembly-fragment"
misc_feature      9451..11335
                  /note="assembly-fragment"
misc_feature      11436..14325
                  /note="assembly-fragment"
misc_feature      14426..17305
                  /note="assembly-fragment"
misc_feature      17406..20166
                  /note="assembly-fragment"
misc_feature      20267..23474
                  /note="assembly-fragment"
misc_feature      23575..26403
                  /note="assembly-fragment"
misc_feature      26504..29899
                  /note="assembly-fragment"
misc_feature      30000..34929
                  /note="assembly-fragment"
misc_feature      35030..41463
                  /note="assembly-fragment"
misc_feature      41564..47890
                  /note="assembly-fragment"
misc_feature      47991..56092
                  /note="assembly-fragment"
misc_feature      56193..63063
                  /note="assembly-fragment"
misc_feature      63164..70819
                  /note="assembly-fragment"
misc_feature      70920..82482
```

	Query Match	Best Local Similarity	2.1%; 67.4%;	Score 188;	DB 2;	Length 149425;
	Matches 507;	Conservative	0;	Mismatches 190;	Indels 55;	Gaps 15;
misc_feature	/note="assembly_fragment"	82583..98645				
misc_feature	/note="assembly_fragment"	98746..117782				
misc_feature	/note="assembly_fragment"	117883..149425				
	/note="assembly_fragment"	clone_end:7				
	vector_side:right"					
BASE COUNT	36100 a	37729 c	37628 g	36165 t	1803 others	
ORIGIN						
Y	3238	AGAGCCAAAGGGCGGCACACGGCAGCTAGAGCTGATAGCTGTGCGGCTAGATT	3237			
Db	44776	AAAGCCAAAGGGTGGGGACA--TCACACTCTCATGTGCAGGCTG--TGGAGGGCTCTGGGCT	44832			
Y	3298	CCGAGGAGACCTGG--GACCTACTTCCACACCCCCACATCCATTTCTCTGAGGGCCCTAT	3355			
Db	44833	CCGAGGAGACCTGGCAGCAGCTGCTCTCCACCCCTCAC-----CTCTCTGGGGCCCTAT	44887			
Y	3356	CTTCCCTTATATGATGTAAGAAATTCCTGGGGGGGGGGGGTGGTGGAGACAAGATC	3415			
Db	44888	CTTCCCTTATATGATGTAAGAAATTCCTAGAGAGAGGGG-----GGTGGGACAAGG--	44940			
Y	3416	GTTGGGCTCCTGACGACGAGCTTGGCACAACTTCCATAGATCTCCAGTGGTGGCTGCC	3475			
Db	44941	--TGGCTCTTCTGAGCAGCACTTGGCCACAATTCCTGAGATCTCCAGTGGGACAGTCC	44998			
Y	3476	TCTTCCAGACAGGTAAAGGCAATTGGGTGGGGACACATGGTGAACCAAGAGTGTGAAGG	3535			
Db	44999	TCCCAAGACAGGTAAAGGCAACCTGGTGGGAACATGGAACCTTAGTGGTGGGAAGG	45058			
Y	3536	GACAGGGTCTT--TGTCTTCTCTGCGAGCTT--GTGCTTCTGTATACACCTTGTATAGT	3593			
Db	45059	GACGGGTCTCTGCTTCTTCTTCGGGCAACTTGGGGCTCTCTGAGCACCTTGGTGAAGGG	45118			
Y	3594	TTGGGGGTGAGATTAAGGTGCTCT--GAAACTCTGAAAAGCAAGAAAGCC--AGCAGGCTG	3650			
Db	45119	TGGAAGGGAGGCTGTACACTCTGATTAACGTGAGAAAGTCTGGGAGCCAGTGTGTCTG	45178			
Y	3651	TCTTGGGCTCTTCAATGAAGAAATTCACAGACCCCTTCTCTGTAA--GTACCTTGGCT	3708			
Db	45179	CCCTGGGGCTCTCCAGGGAAGAACTTTTGGAGGCCCTGCCCCAGACAAATGGCTCTCAGATC	45238			
Y	3709	TCAATCTGTGAATTTCCCTGGGACCAAGTGGCT-----CTGGGAGATCAAT	3757			
Db	45239	TCTACTCTGTGGCTAGGCCCATCCACTTTGGGTGTACCAATCCCAATTTGCCCTGGAACT	45298			
Y	3758	TCTACAAATTAATTCAGGACAGTCCAG-------ACTTGAATCCGTCCTGT	3805			
Db	45299	TCTAGAGTTAAACCAAGACAGTGTGGGACAGCGGAGCTAGTTGGTCTCCGACCTC	45358			
Y	3806	ATTTAC--TACTTCTCTCTGGCTGCTCATTTCTGTGTTCATGTCTTACACATCTGAAT	3862			
Db	45359	ACTGCCCGAGCCCTCCACACAGCTCACTCTTGGCCCGCTCTTATCTGAAGT	45418			
Y	3863	GG--TTCCTTGTGACACATTCCTCTACACTCTCTGGAGGTGTATCTTGGACAT	3919			
Db	45419	GTTTTTTTTCTTTGTGTACACTTCTCTGTACACTCTCTGGACCTCGGGGTCTGTCTCT	45478			
Y	3920	GTAATCTGGAGTGAAGCTGCAGCCACGAGA	3951			
Db	45479	GGCTCTAGCATATGAGCAACGACCCACGAGA	45510			
RESULT 13						
LOCUS	AC021016	191754 bp	DNA	linear	PRI 09-MAY-2001	
DEFINITION	Homo sapiens BAC clone RP11-378A13 from 2, complete sequence.					

ACCESSION AC021016  
 VERSION AC021016.4 GI:11120952  
 KEYWORDS HTG.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 191754)  
 AUTHORS Suiston,J.E. and Waterston,R.  
 JOURNAL Toward a complete human genome sequence  
 MEDLINE Genome Res. 8 (11), 1097-1108 (1998)  
 PUBMED 99063792  
 REFERENCE 2 (bases 1 to 191754)  
 AUTHORS Cordes,M., Maupin,R., Hawkins,M. and Boyer,E.  
 JOURNAL The sequence of Homo sapiens BAC clone RP11-378A13  
 TITLE Unpublished  
 REFERENCE 3 (bases 1 to 191754)  
 AUTHORS Waterston,R.H.  
 JOURNAL Direct Submission  
 TITLE Submitted (12-JAN-2000) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA  
 REFERENCE 4 (bases 1 to 191754)  
 AUTHORS Waterston,R.  
 JOURNAL Direct Submission  
 TITLE Submitted (08-NOV-2000) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
 REFERENCE 5 (bases 1 to 191754)  
 AUTHORS Waterston,R.  
 JOURNAL Direct Submission  
 TITLE Submitted (09-MAY-2001) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
 COMMENT On Nov 8, 2000 this sequence version replaced gi:7630969.  
 ----- Genome Center  
 Center: Washington University Genome Sequencing Center  
 Center code: WUGSC  
 Web site: http://genome.wustl.edu/gsc  
 Contact: saplens@wustl.wustl.edu  
 ----- Summary Statistics  
 ----- Center project name: H\_NH0378A13

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:  
 Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see http://genome.wustl.edu/gsc

SOURCE INFORMATION:  
 The RCI-11 human BAC library was made from the blood of one male donor, as described by Osogawa,K., Woon,P.Y., Zhao,B., Frengen,E., Tatenno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (http://bacpac.med.buffalo.edu)  
 VECTOR: pBACE3.6  
 NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP11-146N10; the clone sequenced to the right is RP11-36O03. Actual start of this clone is at base position 1 of RP11-378A13; actual end is at base position 191754 of RP11-378A13.

The sequence RP11-378A13 from 156426 to 156433 is derived from a single m13 subclone. Size of the region was confirmed by PCR from BAC DNA.

FEATURES	Source	location/Qualifiers
misc_feature	1..191754	/organism="Homo sapiens"
misc_feature	1..224	/mol_type="genomic DNA"
misc_feature	1..224	/db_xref="taxon:9606"
misc_feature	1..224	/chromosome="2"
misc_feature	1..224	/map="2"
misc_feature	1..224	/clone="RP11-378A13"
misc_feature	1..224	/clone_lib="RPCT-11"
misc_feature	1..224	/note="similar to EST AA947005 (NID:g3110400) oq45e10.s1"
misc_feature	1..224	/note="similar to EST A1184556 (NID:g3735194) qd60a10.x1"
misc_feature	1..224	/note="similar to EST A1457236 (NID:g4310105) tk01d06.x1"
misc_feature	1..216	/note="similar to EST AA782751 (NID:g2842082) aj08c02.s1"
misc_feature	1..216	/note="similar to EST AM020425 (NID:g5873955) df09e11.y1"
misc_feature	2..223	/note="similar to EST AA175052 (NID:g1756173) ms82f02.r1"
misc_feature	2..223	/note="similar to EST AV079022 (NID:g5198850)"
misc_feature	2..223	/note="similar to EST AW538472 (NID:g7180898)"
misc_feature	2..49	/note="similar to EST AV330904 (NID:g6370956)"
misc_feature	7..224	/note="similar to EST A1454294 (NID:g4292719)"
misc_feature	7..224	/note="similar to EST AW433991 (NID:g6965298)"
misc_feature	12..223	/note="similar to EST AW059787 (NID:g6652109)"
misc_feature	20..223	/note="similar to EST A1709077 (NID:g4998853) as86d11.x1"
misc_feature	60..220	/note="similar to EST AA870364 (NID:g2965809) vq44e05.r1"
misc_feature	67..223	/note="similar to EST AV330904 (NID:g6370956)"
repeat_region	447..660	/rpt_family="Alu"
repeat_region	1182..1425	/rpt_family="L1"
repeat_region	1442..1583	/rpt_family="MERL-type"
repeat_region	1820..1928	/rpt_family="Alu"
misc_feature	1948..2347	/note="similar to EST AA224009 (NID:g1844621) zr12g06.r1"
repeat_region	2162..2293	/rpt_family="Alu"
repeat_region	2297..2587	/rpt_family="Alu"
repeat_region	2597..2804	/rpt_family="Alu"
misc_feature	2606..2923	/note="similar to EST AW351985 (NID:g6849698)"
misc_feature	2630..3135	/note="similar to EST AA234699 (NID:g1859192) zs39g05.r1"
misc_feature	2804..3242	/note="similar to EST A1916611 (NID:g5636466) wa28f08.x1"
misc_feature	2976..3245	/note="similar to EST A1401669 (NID:g4244756) th24e10.x1"
repeat_region	3097..3284	/rpt_family="MIR"

	misc_feature	3546..3781	/note="similar to EST AA102341 (NID:g1647297) zm26h05.i1"
	repeat_region	3848..4127	/rpt_family="Alu"
	repeat_region	4187..4469	/rpt_family="Alu"
	repeat_region	4474..4690	/rpt_family="Alu"
	misc_feature	4723..5241	/note="similar to EST AW444974 (NID:g6986736)"
	misc_feature	4931..5413	/note="similar to EST AA100813 (NID:g1647248) zm26h05.s1"
	repeat_region	6006..6137	/rpt_family="MIR"
	repeat_region	6243..6394	/rpt_family="MIR"
	misc_feature	6967..7092	/note="similar to EST AW461516 (NID:g7031684)"
	misc_feature	7935..8054	/note="similar to EST AW461516 (NID:g7031684)"
	repeat_region	8294..8433	/rpt_family="MERL_type"
	misc_feature	8553..8604	/note="similar to EST AI391439 (NID:g4217443) tf96e02.x1"
	misc_feature	8556..8749	/note="similar to EST AW461516 (NID:g7031684)"
	misc_feature	9362..9733	/note="similar to EST AI391439 (NID:g4217443) tf96e02.x1"
	misc_feature	9476..9723	/note="similar to EST AA595679 (NID:g2411029) nk66f01.s1"
	misc_feature	9997..10265	/note="similar to EST AI634626 (NID:g4685956) tz30g01.x1"
	misc_feature	10003..10462	/note="similar to EST AI828100 (NID:g5448771) wk31g06.x1"
	misc_feature	10003..10482	/note="similar to EST AI890574 (NID:g5595738) wm87e04.x1"
	misc_feature	10003..10482	/note="similar to EST AI978960 (NID:g5803990) wt68e06.x1"
	misc_feature	10003..10482	/note="similar to EST AL047914 (NID:g4728747)"
	misc_feature	10003..10482	/note="similar to EST AM027349 (NID:g5886105) wf73e09.x1"
	misc_feature	10003..10457	/note="similar to EST AI972200 (NID:g5769026) wv63h01.x1"
	misc_feature	10003..10421	/note="similar to EST W48859 (NID:g1336988) zc42c08.s1"
	misc_feature	10003..10330	/note="similar to EST AM074215 (NID:g6029213) xp09d12.x1"
	misc_feature	10003..10249	/note="similar to EST AA143563 (NID:g1712935) zc32d10.s1"
	misc_feature	10003..10247	/note="similar to EST AI583619 (NID:g4569516) ts16d09.x1"
Query Match                  2.1%; Score 188; DB 9; Length 191754;			
Best Local Similarity    67.4%; Pred. No. 7,9e-43;			
Matches 507; Conservative 0; Mismatches 190; Indels 55; Gaps 15;			
OY	3228	AGAGCGAAGGGGCGGCACACGCACTAGGCTGTACAGGCTTGTCTGGCCTTAGGTT	3297
Dd	164808	AAAGCCAAAGGGGGGCCACA--TCAGCTTCATATTGCAAGCTG-TGAGAGGCTCTGGCT	164864
OY	3298	CCAGAGGACTTG--GCACCTACTTCCCACCACCCCCCATTCATCTCTGGGCCCCATT	3355
Dd	164865	CCGAGGGAACCTBGCACGACACTCCCTCCCAACCCCTCAC-----CTCTGGGGCCCTTA	164919
OY	3356	CTTCCCTTATATGTGAAGAAGTTCCTGGGGGGGGGGTGGTGTGAGAACAAAGCTC	3415
Dd	164920	CTTCCCTTATATGTGAAGAAGGACTTCTGTGAGAGGAGGAGG----GGTGGGACAAAAG-	164972
OY	3416	GTTGGGTCTCTGCAAGCAGCTTGCCACAACTTCTTAAGATCTCCAAGTGTGGCTGCC	3475
Dd	164973	--TGCGCTTCTGCAAGCAGCTTGCCACAAATTCCTGAAGATCTCCAGGTGGAGCTGGCC	165030

OY	3476	TCTTCCAGACAGGTAAGAACAATTGGGTGGGACACATGGTCACACAGGTTGTTGAAGG	3535
Db	165031	TCGCCAAGACAGGTAAGCAACCTGTGGGAAACATGAGAGACTTAGTGTTGCAGAGG	1650
OY	3536	GACAGGGTCCCT-TGCTTCTCTCTCGGACGCTT-GTGTCTTCTGTATGGACCTTGGTATAGT	3593
Db	165091	GACGGGGTCCCTCTCTCTTTCTCGGGCAAATTGGGGCCTCTCTGGGACACTTGTATACGGG	16511
OY	3594	TTGGGGGTGAGGTAAAGTGCTCT-GAAACTCTGAAGAACAGCAAGACC--AGCAGGCTG	3650
Db	165151	TCGAAGGGGAGGCTTGACACTCTGGATTAGCSTGAAAGTGTGTGGAGGCCCATGTGTCTG	1652
OY	3651	TCTTGGGCTTCAATGAAAGAGTTCAACAGACCCCCTTCTCTGTAA--GTACACCTTCGT	3708
Db	165211	CCCTGGGGCTCCAGGAGAGACTTTTGAAGGCCCTGGCCCCAGACAATAGGCTCTCTCAGTC	1652
OY	3709	TCATCTGTATAGATTCCCTGGGACCAAGTGTGCT-----CCTGGGACTCAGATT	3757
Db	165271	TCCATACCTGTGGTACAGCCCCCATCTCCACTTTGGGTGTACACATCCCAATTCCTCGAATC	1653
OY	3758	TCTCAATTTAAATACAGAGACAGTCTCTAG-----ACTTGGACCTCGGCGCTGT	3805
Db	165331	TCTAGAGTTAAACACAGAGACAGTCTGGGACAGCCGAGACTAGTTGGTCTTCCCTGACCTC	1653
OY	3806	ATTAC--TACTTCTCTGTGGCTGCTCATTTCTGTTCATGTCTTACACATCTGAAAT	3862
Db	165391	ACTCCCGAGGCCCTCCACACAGCTCATCTCTTTGCCGTGGTTCTTCACTGTGAAT	1654
OY	3863	GG---TTTTCTTGTGTACACATTCCTCGACACTCTGTGGAGGTCTATCTTGGGCAT	3919
Db	165451	GTTTTTTCTTTGTGTGATCACCACTTCTCTGTGACACTCTGGAGACTCCGGGTCTGCTCTCT	1655
OY	3920	GTATCCGTGGATGTAACTGCAGCCACACAGA	3951
Db	165511	GGCTCCTAGCATATGACCAACAGCCACACAGA	165542
<hr/>			
RESULT 14			
<hr/>			
LOCUS	G76462	568 bp DNA linear	STS 06-SEP-2002
DEFINITION	S208Pb543PH7.T0 129S1/SvImJ Mus musculus STS genomic, sequence tagged site.		
ACCESSION	G76462		
VERSION	G76462.1 GI:22727218		
KEYWORDS	STS.		
SOURCE	Mus musculus (house mouse)		
ORGANISM	Mus musculus		
REFERENCE	Eukaryota; Metazoa; Chordata; Cranialta; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
AUTHORS	Wade,C.		
TITLE	Polymorphism Structure in the Mouse		
JOURNAL	Unpublished (2002)		
COMMENT			

Contact: Kerstin Lindblad-Toh  
Whitehead Institute for Biomedical Research, Center for Genome  
Research  
320 Charles Street, Cambridge, MA 02141, USA  
Tel: 6172521477  
Fax: 6172580903  
Email: kersli@genome.wi.mit.edu  
Primer A: None  
Primer B: None  
STS size: 568  
Protocol:  
WGS-disccovery: Paired-end low-coverage whole genome shotgun reads  
were generated from 125S1/SVintJ, C3H/HeJ, and BALB/cBYJ. The WGS  
reads were placed uniquely on the MGS/CV3 C57BL/6J assembly and SNP  
detection was carried out by SSAHA-SNP. 225,000 reads were  
annotated  
as STS and 81,000 SNPs were annotated with alleles from C57BL/6J  
and the strain from which the particular read came. The validation  
rate for these SNPs was estimated at approximately 98%.

FEATURES	source	location/Qualifiers
		1..568
		/organism="Mus musculus"
		/mol_type="genomic DNA"
		/strain="129S1/SvImJ"
		/db_xref="taxon:10090"
		/map_ref=1 22-501 7012032-7012511"
		/clone_lib="129S1/SvImJ"
		<1..>568
STS		
BASE COUNT	176 a 122 c 111 g 159 t	
ORIGIN		
Query Match	2.1%; Score 186.6; DB 11; Length 568;	
Best Local Similarity	69.2%; Pred. No. 6.5e-43;	
Matches 351; Conservative	0; Mismatches 129; Indels 27; Gaps 6;	
QY	1847 TATGTGGGAGGAGAGATGGCTCAGCTTCACAGAGACACTGGTCGTCTTGACAGAGACC	1906
Db	79 TGTATTGCTGTGGTTTAAAGACTCAGTGTCMAAGACACTGGCTCTTCCAGAGGACC	138
QY	1907 TAGATTCAAGTCCAGAGCACTATFATMGGGCTCAGACAGCACTCTGTAATCCAGTCCGA	1966
Db	139 AGGGCTCAGTTCTCATPAAGCCAGAGATGGCTCACACTATTTGTACTCCAGTCCGAG	198
QY	1967 GGGTTCCACACCCCTCTTCTGGCCTCCACAGGACCACATPACA-TAGTACAGACATACA	2025
Db	199 GCATCTGGTGCCCTCTCTCTGGGCTTCACAGGAATTTCTATPACAAGTACACAGATCA	258
QY	2026 TGCAGGCAAAACCCATACACATATAATTAATGAAGAACTTAAAGGTGCATGCTT	2085
Db	259 TGC-----AAACACTCATACACATPAAATCAATTCATTTAAATTTACACATAAAT	313
QY	2086 GGTAAACATTTGTGCTTACACATGCTGATTTGAAGACATGTACACAGCACACATGAAGAG	2145
Db	314 GAT-AACAAAGCTGTCCAGAAATTTAGATAATATATGCTTAAATTTCTG-----	366
QY	2146 GATCTGGGGCTGGAGAGATGGCTCAGCGGTTAAGACACACTACTGCTTTCCGAAGCAG	2205
Db	367 --GTAGGGGGCTGGAGTGAATGGCTCAGTGAAGAGCACTACGCTCTTCC---GAAG	420
QY	2206 GTCCGTGAGTCAAAATCCTAGACACACATGGTGGCTCAACCAATCATTAATGAGATCTG	2265
Db	421 GTCCGTGAGTCAAAATCCAGACACACATGGTGGCTCAACCAATCATTAATGAGATCTG	480
QY	2266 ACACCCCTCTTGTGTGTCATCTGAAGACAGCTGCACAGAGCTACAGTACTTAATPACTA	2325
Db	481 ACTCCCTCTTCTGGTGTGTCTGAAGAC-----AGCTACAGTCTACTCATATATAATA	532
QY	2326 ATAAATTAATCTTTTAAATAATG 2352	
Db	533 AATTAATCTTTTAAATGTTG 559	
RESULT 15		
AC100085/c	195636 bp DNA linear HTG 23-APR-2003	
LOCUS	Mus musculus clone RP23-36K2, WORKING DRAFT SEQUENCE, 6 unordered	
DEFINITION	pieces.	
ACCESSION	AC100085	
VERSION	AC100085.3 GI:30018196	
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT.	
SOURCE	Mus musculus (house mouse)	
ORGANISM	Mus musculus	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
AUTHORS	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.	
TITLE	1 (bases 1 to 195636)	
JOURNAL	Birren,B., Nusbaum,C. and Lander,E.	
REFERENCE	Mus musculus, clone RP23-36K2	
AUTHORS	Unpublished	
	2 (bases 1 to 195636)	
	Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,	
	Anderson,S., Barina,N., Bastien,V., Boguslavsky,L., Boukhalter,B.,	
	Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,	

Choepeel, Y., Collangelo, M., Collins, S., Collimore, A., Cook, A.,  
Cooke, P., DeArrellano, K., Dewar, K., Diaz, J.S., Dodge, S., Faro, S.,  
Ferreira, P., Fitzhugh, M., Gage, D., Galagan, J., Gairyna, S.,  
Hagios, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,  
Hagos, B., Heaford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R.,  
Jones, C., Kamat, A., Karatas, A., Kells, C., LaRoque, K.,  
Lamas, R., Landers, T., Lehoczy, J., Levine, R., Liu, G.,  
Maclean, C., MacDonald, P., Major, J., Margulis, N., Matthews, C.,  
McCarthy, M., McEwan, P., McKernan, K., McPheters, R., Meldrum, J.,  
Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C.,  
Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D.,  
Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V.,  
Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P.,  
Roman, J., Rosetti, R., Roy, A., Santos, R., Schauer, S., Schnappack, R.,  
Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,  
Strauss, N., Sudramanlan, A., Talamas, J., Tesfaye, S., Theodore, J.,  
Topham, K., Travers, M., Travis, N., Triggillo, V., Vassiliev, H.,  
Viel, R., Vo, A., Wilson, B., Wu, X., Wymen, D., Ye, W.J., Young, G.,  
Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission  
Submitted (22-NOV-2001) Whitehead Institute/MIT Center for Genome  
Research, 330 Charles Street, Cambridge, MA 02141, USA  
3 (bases 1 to 195636)

Bitren, B., Nusbaum, C., Lander, E., Abouelleil, A., Allen, N.,  
Anderson, S., Atachchi, H.M., Barna, N., Bastien, V., Bloom, T.,  
Boguslavsky, L., Boukhalter, B., Camarata, J., Chang, J., Choepeel, Y.,  
Collimore, A., Cook, A., Cooke, P., Corum, B., DeArrellano, K.,  
Diaz, J.S., Dodge, S., Dooley, K., Dorris, L., Erickson, J., Faro, S.,  
Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J., Gairyna, S.,  
Graham, L., Grand-Pierre, N., Halez, N., Hagopian, D., Hagos, B.,  
Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,  
Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R.,  
Lindblad-Toh, K., Liu, G., Liu, A., Mabbitt, R., Maclean, C.,  
MacDonald, P., Major, J., Manning, J., Matthews, C., McCarthy, M.,  
Meldrum, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J.,  
Nguyen, C., Nicol, R., Norbu, C., O'Connor, T., O'Donnell, P.,  
O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N.,  
Rachpaka, A., Ramasamy, U., Raymond, C., Retta, R., Rise, C., Rogov, P.,  
Roman, J., Schauer, S., Schnappack, R., Seaman, S., Severy, P., Smith, C.,  
Spencer, B., Stange-Thomann, N., Stojanovic, N., Stubbs, M.,  
Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M.,  
Vassiliev, H., Venkataraman, V.S., Viel, R., Vo, A., Wilson, B., Wu, X.,  
Wymen, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission  
Submitted (23-APR-2003) Whitehead Institute/MIT Center for Genomes  
Research, 330 Charles Street, Cambridge, MA 02141, USA  
On Apr 17, 2003 this sequence version record replaced gi:22004247.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR  
Web site: <http://www-seq.wi.mit.edu>  
Contact: sequence.submissions@genome.wi.mit.edu

Project Information  
Center project name: LI3938  
Center clone name: 36\_K\_2

NOTE: This is a 'working draft' sequence. It currently  
consists of 6 contigs. The true order of the pieces  
is not known and their order in this sequence record is  
arbitrary. Gaps between the contigs are represented as  
runs of N, but the exact sizes of the gaps are unknown.  
This record will be updated with the finished sequence  
as soon as it is available and the accession number will  
be preserved.

1 1071: contig of 1071 bp in length  
\* 1072 1171: gap of 100 bp  
\* 1172 3389: contig of 2228 bp in length  
\* 3400 3499: gap of 100 bp  
\* 3500 18021: contig of 14522 bp in length  
\* 18022 18121: gap of 100 bp







GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 13, 2003, 12:10:01 ; Search time 15870 Seconds  
(without alignments)  
13775.599 Million cell updates/sec

Title: US-09-877-935-1  
Perfect score: 8995  
Sequence: 1 gatctgctgcacccaagaca.....ctctagctcgcacacatg 8995

Scoring table: OLIGO NUC  
Gapop 60.0 , Gapext 60.0

Searched: 22781392 seqs, 12152238056 residues

Word size : 0

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

EST:\*  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estnu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hic:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_hic:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: em\_gss\_hum:\*  
18: em\_gss\_inv:\*  
19: em\_gss\_pin:\*  
20: em\_gss\_vrt:\*  
21: em\_gss\_fun:\*  
22: em\_gss\_mam:\*  
23: em\_gss\_mus:\*  
24: em\_gss\_pro:\*  
25: em\_gss\_rtd:\*  
26: em\_gss\_phg:\*  
27: em\_gss\_vrl:\*  
28: gb\_gss1:\*  
29: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	498	5.5	507	28	AZ354198 1M0093E23
2	322	3.6	336	28	AZ657211 1M0533G15
3	262	2.9	307	28	AZ354559 1M0093E23
4	242	2.7	291	28	AZ821333 2M0094K06

5	206	2.3	495	28	BH086621 RPT-24-2
6	137	1.5	639	28	AZ821403 2M0094I08
7	81	0.9	565	28	AZ747682 RPT-24-8
8	80	0.9	373	10	BB813805 BB813805
9	79	0.9	713	28	AZ902196 RPT-24-1
10	76	0.8	671	10	BB662563 BB662563
11	76	0.8	1083	11	AK086147 Mus muscu
12	75	0.8	438	14	CD538197 B0201B07-
13	75	0.8	540	28	AZ427733 1M0209H24
14	74	0.8	674	28	CNS02J2Y AL200887 Tetraodon
15	71	0.8	779	13	BU705960 UI-M-FRO-
16	70	0.8	321	12	BM219657 C0928C02-
17	67	0.7	499	28	BZ134506 C8230-385
18	67	0.7	525	28	BH078848 RPT-24-2
19	67	0.7	676	29	DR4814T AL980584 Danio rer
20	67	0.7	705	29	BX184553 Danio rer
21	67	0.7	713	29	BX154096 Danio rer
22	67	0.7	722	29	BX133701 Danio rer
23	67	0.7	726	13	BX076218 BX076218
24	66	0.7	138	29	BX217769 Danio rer
25	66	0.7	167	9	AA078152 CA489703 omykrtb0
26	66	0.7	214	14	CA889703 omykrtb0
27	66	0.7	296	14	CA302665 faa15b02.
28	66	0.7	302	10	BB347599 BB347599
29	66	0.7	371	28	BH125739 RPT-24-3
30	66	0.7	377	29	BX239417 Danio rer
31	66	0.7	408	9	AI032142 AI032142 os76d08.s
32	66	0.7	410	28	BH398554 AG-ND-148
33	66	0.7	448	29	BX165815 Danio rer
34	66	0.7	460	10	BE606136 f123c06.Y
35	66	0.7	507	10	BF468206 UI-M-CCO-
36	66	0.7	567	29	DR5M24S AL747210 Danio rer
37	66	0.7	573	29	AL742982 Danio rer
38	66	0.7	574	28	AZ833285 2M0115G17
39	66	0.7	584	28	AZ412557 AZ412557 1M0186N01
40	66	0.7	586	29	BX184359 Danio rer
41	66	0.7	594	29	BX165614 Danio rer
42	66	0.7	612	12	BI839570 fg39h09.Y
43	66	0.7	620	29	BX152328 Danio rer
44	66	0.7	625	29	U68740 U68740 Rftt
45	66	0.7	667	29	BX159688 Danio rer

#### ALIGNMENTS

RESULT 1  
AZ354198/c 507 bp DNA linear GSS 02-OCT-2000  
LOCUS  
DEFINITION  
1M0093E23F Mouse 10kb plasmid UGCGIM library Mus musculus genomic  
clone UGCGIM0093E23 F, genomic survey sequence.

ACCESSION  
AZ354198  
VERSION  
AZ354198.1 GI:10465402

KEYWORDS  
GSS.  
Mus musculus (house mouse)

ORGANISM  
Mus musculus  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE  
1 (bases 1 to 507)  
Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,  
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly  
M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.  
and Wright,D., Weis,R.  
Mouse whole genome scaffolding with paired end reads from 10kb  
plasmid inserts

TITLE  
Unpublished

JOURNAL  
COMMENT  
Contact: Robert B. Weiss  
University of Utah Genome Center  
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SIC, UT  
84112, USA  
Tel: 801 585 5606  
Fax: 801 585 7177

Email: ddunn@genetics.utah.edu  
Insert Length: 10000 Std Error: 0.00  
Plate: 0093 row: E column: 23  
Seq primer: CGTTGTAACGACGCGCCAGT  
Class: Plasmid ends  
High quality sequence stop: 507.

## FEATURES

source

Location/Qualifiers

1. 507  
/organism="Mus musculus"  
/mol\_type="genomic DNA"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="UUGC1M0093E23"  
/sex="Male"  
/lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"  
/clone\_lib="Mouse 10kb plasmid UUGC1M library"  
/note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource  
(http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adapted DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of PWD42 (g114732114[gb|AF129072.1]), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adapted mouse DNA was annealed to adapted vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance.

## BASE COUNT

125 a 139 c 105 g 138 t

## ORIGIN

## Query Match

Best Local Similarity 5.5%; Score 498; DB 28; Length 507;

Matches 498; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

8431 CATCCAGAGTGGTACGAGGAGACTCAGGATATGCTGCTCCCTCCACAGTGT 8490  
|||||  
507 CATCCAGAGTGGTACGAGGAGACTCAGGATATGCTGCTCCCTCCACAGTGT 448  
8491 AGAATTACAGGCATATCACTGCTGGAAGATTTTAACTCCTGAGATAGACA 8550  
|||||  
447 AGAATTACAGGCATATCACTGCTGGAAGATTTTAACTCCTGAGATAGACA 388  
8551 GGCACTTACCAATGAGAGGTTCTTTTGTGTTGGTTTCCCTGATTAAGTCA 8610  
|||||  
387 GGCACTTACCAATGAGAGGTTCTTTTGTGTTGGTTTCCCTGATTAAGTCA 328  
8611 GGCACTTGAAGTGTAGAGTGGCTTACATCACTTGTCTCAAAAAGCCTATAGAG 8670  
|||||  
327 GGCACTTGAAGTGTAGAGTGGCTTACATCACTTGTCTCAAAAAGCCTATAGAG 268  
8671 GTAGGAGTGTAGAGTGGCTTAAAGAGCCGCTGTGATAGCACACAGATAGCC 8730  
|||||  
267 GTAGGAGTGTAGAGTGGCTTAAAGAGCCGCTGTGATAGCACACAGATAGCC 208  
8731 TGCACTATATAGCAAGACCTGTTTCAAAAACATGAGAGGAGGTTATGTTAAGTCT 8790  
|||||  
207 TGCACTATATAGCAAGACCTGTTTCAAAAACATGAGAGGAGGTTATGTTAAGTCT 148  
8791 GGCGTGTGTACAGGACCTAAGGACCAATGTAGACTTGTGACTAAGAAAGATCATCA 8850  
|||||  
147 GGCGTGTGTACAGGACCTAAGGACCAATGTAGACTTGTGACTAAGAAAGATCATCA 88  
8851 TCAAAGCCGGGTGGGACAGGTGAGAGTGTGACTACAGTGTCAAGACCCCATAGGAAGC 8910  
|||||  
87 TCAAAGCCGGGTGGGACAGGTGAGAGTGTGACTACAGTGTCAAGACCCCATAGGAAGC 28

QY 8911 CAGTTCCCTCTCTCTC 8928  
|||||  
DB 27 CAGTTCCCTCTCTCTC 10

## RESULT 2

AZ657211 336 bp DNA GSS 14-DEC-2000  
LOCUS AZ657211/c  
DEFINITION 1M0533G15F Mouse 10kb plasmid UUGC1M library Mus musculus genomic  
clone UUGC1M0533G15 F, genomic survey sequence.

ACCESSION AZ657211  
VERSION AZ657211.1 GI:11794357  
KEYWORDS GSS.

## SOURCE

## ORGANISM

Mus musculus (house mouse)

## REFERENCE

AUTHORS

## TITLE

## JOURNAL

## COMMENT

Unpublished  
Contact: Robert B. Weiss  
University of Utah Genome Center  
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SIC, UT  
84112, USA  
Tel: 801 585 5606  
Fax: 801 585 7177  
Email: ddunn@genetics.utah.edu  
Insert length: 10000 Std Error: 0.00  
Plate: 0533 row: G column: 15  
Seq primer: CGTTGTAACGACGCGCCACT  
Class: plasmid ends  
High quality sequence stop: 336.

## FEATURES

source

Location/Qualifiers

1. 336  
/organism="Mus musculus"  
/mol\_type="genomic DNA"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="UUGC1M0533G15"  
/sex="Male"  
/lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"  
/clone\_lib="Mouse 10kb plasmid UUGC1M library"  
/note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource  
(http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adapted DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of PWD42 (g114732114[gb|AF129072.1]), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adapted mouse DNA was annealed to adapted vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance.

## BASE COUNT

103 a 88 c 66 g 79 t

## ORIGIN

Query Match 3.6%; Score 322; DB 28; Length 336;  
Best Local Similarity 100.0%; Pred. No. 2e-102;

Matches 322; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5431 GCAAGCCTTTGTCCTCCCTGCTGTCATATAACACCCCTGTCATACATTAGTTAT 5490  
 |||||  
 Db 336 GCAAGCCTTTGTCCTCCCTGCTGTCATATAACACCCCTGTCATACATTAGTTAT 277  
 |||||  
 QY 5491 TTTACTGTCAGTTTGTCTCCAGACAGTCATCTGTAGACCTCTGCTCTTAACACCA 5550  
 |||||  
 Db 276 TTACGTCAGTTTGTCTCCAGACAGTCATCTGTAGACCTCTGCTCTTAACACCA 217  
 |||||  
 QY 5551 GGTATGCCCCACATCTCTACCCAGACAGTCAGACAGCCTTACAGAAAGGTATAC 5610  
 |||||  
 Db 216 GGTATGCCCCACATCTCTACCCAGACAGTCAGACAGCCTTACAGAAAGGTATAC 157  
 |||||  
 QY 5611 AGTACCAAGATGGCCAGATATAACAAACCTACATCTTGTACCAATTTGTTT 5670  
 |||||  
 Db 156 AGTACCAAGATGGCCAGATATAACAAACCTACATCTTGTACCAATTTGTTT 97  
 |||||  
 QY 5671 GCTGAACGAGAGGGGGTGTGAGTATATGTGTGTGTGTGTGTGTGTGTGTGTGT 5730  
 |||||  
 Db 96 GCTGAACGAGAGGGGGTGTGAGTATATGTGTGTGTGTGTGTGTGTGTGTGTGT 37  
 |||||  
 QY 5731 GT 5752  
 |||||  
 Db 36 GT 15  
 |||||

RESULT 3  
 AZ354559

LOCUS 307 bp DNA linear GSS 02-OCT-2000  
 DEFINITION 1M0093E23R Mouse 10kb plasmid UGCGIM library Mus musculus genomic  
 accession AZ354559  
 version AZ354559.1 GI:10466121  
 keywords GSS.

SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 307)  
 Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamli, C.,  
 Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly,  
 M., Rose, M., Rose, R., Stokes, R., Tinney, A., von Niederhausern, A.  
 and Wright, D., Weis, R.

TITLE Mouse whole genome scaffolding with paired end reads from 10kb  
 plasmid inserts

JOURNAL Unpublished  
 COMMENT Contact: Robert B. Weiss  
 University of Utah Genome Center  
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT  
 84112, USA  
 Tel: 801 585 5606  
 Fax: 801 585 7177

FEATURES  
 source 1..307  
 /organism="Mus musculus"  
 /mol.type="genomic DNA"  
 /strain="C57BL/6J"  
 /db.xref="taxon:10090"  
 /clone="UGCGIM0093E23"  
 /sex="Male"  
 /lab.host="E. Coli strain XL10-Gold, TI-resistant, F-"  
 /clone.lib="Mouse 10kb plasmid UGCGIM library"  
 /note="Vector: PMD42ny; Purified genomic DNA from M.  
 musculus C57BL/6J (male) was obtained from the Jackson  
 Laboratory Mouse DNA Resource  
 High quality sequence stop: 307.  
 Location/Qualifiers

(http://www.jax.org/resources/documents/dnares/). The DNA  
 was hydrodynamically sheared by repeated passage through a  
 0.005 inch orifice at constant velocity. The sheared DNA  
 was blunt end-repaired with T4 DNA polymerase and T4  
 polynucleotide kinase. Adaptor oligonucleotides were  
 ligated to the blunt ends in high molar excess. The  
 adaptor DNA was purified and size-selected for a 9.5 to  
 10.5 kb range using preparative agarose gel  
 electrophoresis. Vector DNA was prepared from a derivative  
 of pMD42 (g114732119b1AF129072.1), a copy-number  
 inducible derivative of plasmid R1. The vector was ligated  
 with adaptors complementary to the insert adaptors and  
 purified. The sheared, adaptor mouse DNA was annealed to  
 adaptor vector DNA, and transformed into  
 chemically-competent E. coli XL10-Gold (Stratagene) cells  
 and selected for ampicillin resistance."

BASE COUNT 63 a 67 c 83 g 94 t

Query Match 2.9%; Score 262; DB 28; Length 307;  
 Best Local Similarity 100.0%; Pred. No. 2.7e-81;  
 Matches 262; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1128 TGGACACAGCTCTCTAGGCTATCTTATTTTAAATATATATTCACAC 1187  
 |||||  
 Db 46 TGGACACAGCTCTCTAGGCTATCTTATTTTAAATATATATTCACAC 105  
 |||||

QY 1188 CGGCTGTGTGGCAGACAGCCTTTATCCAGACCTTGAAGCTGAGCTGAGATTAT 1247  
 |||||  
 Db 106 CGGCTGTGTGGCAGACAGCCTTTATCCAGACCTTGAAGCTGAGCTGAGATTAT 165  
 |||||

QY 1248 ACACACAGGCGAGCTGGGGTGCAGAGCTGGCCGCTTTTGTGTTTCTTTATATG 1307  
 |||||  
 Db 166 ACACACAGGCGAGCTGGGGTGCAGAGCTGGCCGCTTTTGTGTTTCTTTATATG 225  
 |||||

QY 1308 CACTGTGTCTTACCTGCTGTATGTCCGTGCAAGGCTCAGATCCCTTGAGCTGAG 1367  
 |||||  
 Db 226 CACTGTGTCTTACCTGCTGTATGTCCGTGCAAGGCTCAGATCCCTTGAGCTGAG 285  
 |||||

QY 1368 TTAAGACAGTTGTATACGC 1389  
 |||||  
 Db 286 TTAAGACAGTTGTATACGC 307  
 |||||

RESULT 4  
 AZ821333/C

LOCUS 291 bp DNA linear GSS 20-FEB-2001  
 DEFINITION 2M0094R06F Mouse 10kb plasmid UGCGIM library Mus musculus genomic  
 accession AZ821333  
 version AZ821333.1 GI:12991241  
 keywords GSS.

SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 291)  
 Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamli, C.,  
 Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly,  
 M., Rose, M., Rose, R., Stokes, R., Tinney, A., von Niederhausern, A.  
 and Wright, D., Weis, R.

TITLE Mouse whole genome scaffolding with paired end reads from 10kb  
 plasmid inserts

JOURNAL Unpublished  
 COMMENT Contact: Robert B. Weiss  
 University of Utah Genome Center  
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT  
 84112, USA  
 Tel: 801 585 5606  
 Fax: 801 585 7177  
 Email: ddunn@genetics.utah.edu  
 Insert length: 10000 Std Error: 0.00



REFERENCE 1 (bases 1 to 639)  
 AUTHORS Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C.,  
 Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly,  
 M., Rose, M., Rose, R., Stokes, R., Tingy, A., von Niederhausern, A.  
 and Wright, D., Weiss, R.  
 TITLE Mouse whole genome scaffolding with paired end reads from 10kb  
 plasmid inserts  
 JOURNAL Unpublished  
 COMMENT Contact: Robert B. Weiss  
 University of Utah Genome Center  
 University of Utah  
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT  
 84112, USA  
 Tel: 801 585 5606  
 Fax: 801 585 7177  
 Email: ddunn@genetics.utah.edu  
 Insert length: 10000 Std Error: 0.00  
 Plate: 0094 row: 1 column: 08  
 Seq primer: CGTTGTAAACGACGGCCACT  
 Class: plasmid ends  
 High quality sequence stop: 639.

FEATURES  
 source  
 1.639  
 /organism="Mus musculus"  
 /mol\_type="genomic DNA"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="UUGC2M0094108"  
 /sex="Male"  
 /lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"  
 /clone\_lib="Mouse 10kb plasmid UUGC1M library"  
 /note="Vector: PWD42nv; Purified genomic DNA from M.  
 musculus C57BL/6J (male) was obtained from the Jackson  
 Laboratory Mouse DNA Resource  
 (http://www.jax.org/resources/documents/dnares/). The DNA  
 was hydrodynamically sheared by repeated passage through a  
 0.005 inch orifice at constant velocity. The sheared DNA  
 was blunt end-repaired with T4 DNA polymerase and T4  
 polynucleotide kinase. Adaptor oligonucleotides were  
 ligated to the blunt ends in high molar excess. The  
 adaptor DNA was purified and size-selected for a 9.5 to  
 10.5 kb range using preparative agarose gel  
 electrophoresis. Vector DNA was prepared from a derivative  
 of PWD42 (g114732114|gb|AF129072.1), a copy-number  
 inducible derivative of plasmid R1. The vector was ligated  
 with adaptors complementary to the insert adaptors and  
 purified. The sheared, adaptor mouse DNA was annealed to  
 adaptor vector DNA, and transformed into  
 chemically-competent E. coli XL10-Gold (Stratagene) cells  
 and selected for ampicillin resistance."

BASE COUNT 163 a 166 c 121 g 189 t  
 ORIGIN

Query Match 1.5%; Score 137; DB 28; Length 639;  
 Best Local Similarity 100.0%; Pred. No. 2e-37;  
 Matches 137; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2694 CGATTCAGAGTGTGCAAGATCTACAGGGGCGAGAGGGGGAAGAGGCCAT 2753  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 179 CGATTCAGAGTGTGCGAAGATCTACAGGGGCGAGAGGGGGAAGAGGCCAT 120  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||  
 QY 2754 GGTTCACAGACCTTACAGCAGAGGCGAGAGGAGTCCAGGCGCAGGAG 2813  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 119 GGTTCACAGACCTTACAGCAGAGGCGAGAGGAGTCCAGGCGCAGGAG 60  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||  
 QY 2814 GTGAGGCCCTTGTTC 2830  
 ||||||||||||||||||  
 Db 59 GTGAGGCCCTTGTTC 43  
 ||||||||||||||||

RESULT 7  
 A2747682 565 bp DNA linear GSS 25-JAN-2001  
 LOCUS

DEFINITION RPCI-24-80N12.TJB RPCI-24 Mus musculus genomic clone RPCI-24-80N12,  
 genomic survey sequence.  
 ACCESSION A2747682  
 VERSION A2747682.1 GI:12531762  
 KEYWORDS GSS.  
 SOURCE  
 ORGANISM Mus musculus (house mouse)  
 Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 565)  
 AUTHORS Zhao, S., Niernan, W., Malek, J., Shatsman, S., Akinret, B., Levins, M.,  
 Tsegaye, G., Geer, K., Krol, M., Shvartsbeyn, A., Gebregeorgis, E.,  
 Russell, D., de Jong, P., and Fraser, C.M.  
 TITLE Mouse BAC End Sequences from Library RPCI-24  
 JOURNAL Unpublished  
 COMMENT Contact: Shaying Zhao  
 Department of Eukaryotic Genomics  
 The Institute for Genomic Research  
 9712 Medical Center Dr., Rockville, MD 20850, USA  
 Tel: 301 838 0200  
 Fax: 301 838 0208  
 Email: szhao@tigr.org  
 Clones are derived from the mouse BAC library RPCI-24. For BAC  
 library availability, please contact Pieter de Jong  
 (pdejong@mail.cho.org). Clones may be purchased from BACPAC  
 Resources (http://www.choi.org/bacpac/orderingframe.htm). BAC end  
 page: http://www.tigr.org/tdb/bac\_ends/mouse/bac\_end\_intro.html  
 Plate: 80 row: N column: 12  
 Seq primer: SP6  
 Class: BAC ends.

FEATURES  
 source  
 1.565  
 /organism="Mus musculus"  
 /mol\_type="genomic DNA"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="RPCI-24-80N12"  
 /sex="Male"  
 /cell\_type="Spleen/Brain"  
 /clone\_lib="RPCI-24"  
 /note="Vector: pFARBAC1; Site\_1: BamHI; Site\_2: BamHI;  
 RPCI-24 Mouse BAC library produced by Pieter de Jong. The  
 library was cloned in the pFARBAC1 cloning vector at the  
 BamHI sites using MBOI partially digested male C57BL/6J  
 DNA."

BASE COUNT 217 a 119 c 106 g 123 t  
 ORIGIN

Query Match 0.3%; Score 81; DB 28; Length 565;  
 Best Local Similarity 100.0%; Pred. No. 1.e-17;  
 Matches 81; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4593 GCAGTGTGCGACACGCCCTTTAATCCAGCAGCTTGGAGGACAGACAGATTTCG 4652  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 83 GCAGTGTGCGACACGCCCTTTAATCCAGCAGCTTGGAGGACAGACAGATTTCG 142  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||  
 QY 4653 AGTTCAAGCGCCAGCCTGGTCT 4673  
 ||||||||||||||||||  
 Db 143 AGTTCAAGCGCCAGCCTGGTCT 163  
 ||||||||||||||||

RESULT 8  
 BB813805/c 373 bp mRNA linear EST 19-NOV-2001  
 LOCUS BB813805  
 DEFINITION BB813805 RIKEN full-length enriched, lung RCB-0558 LLC cDNA Mus  
 musculus cDNA clone G730020N16 3', mRNA sequence.  
 ACCESSION BB813805  
 VERSION BB813805.1 GI:16986434  
 KEYWORDS EST.  
 SOURCE  
 ORGANISM Mus musculus (house mouse)  
 Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 373)  
 AUTHORS Akimura,T., Atakawa,T., Carninci,P., Furuno,M., Hanagaki,T., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Imocani,K., Ishii,Y., Ito,M., Kawai,J., Kojima,Y., Kono,H., Kouda,M., Matsuyama,T., Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Okazaki,Y., Okido,T., Saito,R., Sakai,K., Sakai,K., Sakazume,N., Sasaki,D., Sato,K., Shibata,K., Shingawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagawa,A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toya,T., Watabiki,A., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.  
 RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al. 2001)

JOURNAL Unpublished  
 COMMENT Contact: Yoshihide Hayashizaki  
 Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute  
 The Institute of Physical and Chemical Research (RIKEN)  
 1-7-22 Suenro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
 Tel: 81-45-503-9222  
 Fax: 81-45-503-9216  
 Email: genome-res@gscl.riken.go.jp,  
 URL:http://genome.gsc.riken.go.jp/  
 Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Kono,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.  
 Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)  
 wagi,K., Fujiwaki,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watabiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsunura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.  
 RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)  
 Kono,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara,Y. and Hayashizaki,Y.  
 Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
 Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.  
 e mouse tissues.

FEATURES  
 source Location/Qualifiers  
 1..373  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:10090"  
 /clone="G730020N16"  
 /tissue\_type="lung"  
 /cell\_line="RCB-0558 LLC"  
 /clone\_lib="RIKEN full-length enriched, lung RCB-0558 LLC cDNA"

BASE COUNT  
 ORIGIN 99 a 83 c 72 g 119 t

Query Match 0.9%: Score 80; DB 10; Length 373;  
 Best Local Similarity 100.0%; Pred. NO. 2.7e-17;  
 Matches 80; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2202 GAAGGCTGCTGAGTCAATCTTACCAACCAATGCTGCTACCAACCATCATATGAGA 2261  
 |||||||  
 DB 140 GAAGCTGCTGAGTCAATCTTACCAACCAATGCTGCTACCAACCATCATATGAGA 81  
 |||||||

OY 2262 TCTGACACCCCTCTTCTGTG 2281  
 |||||||  
 DB 80 TCTGACACCCCTCTTCTGTG 61  
 |||||||

RESULT 9 713 bp DNA linear GSS 05-MAR-2001  
 A2902196  
 LOCUS RPCI-24-158J11.TU RPCI-24 Mus musculus genomic clone RPCI-24-158J11  
 DEFINITION , genomic survey sequence.  
 ACCESSION A2902196

VERSION A2902196.1 GI:13221141  
 KEYWORDS GSS.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 Mammalia: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi: Mammalia: Eutheria: Rodentia: Sciurognathi: Muridae: Murinae: Mus.  
 1 (bases 1 to 713)  
 AUTHORS Zhao,S., Nierman,W., Malek,J., Shatsman,S., Akintre,B., Levins,M., Tseaye,G., Geer,K., Krol,M., Shvartsbeyn,A., Gebregeorgis,E., Russell,D., de Jong,P. and Fraser,C.M.  
 Mouse BAC End Sequences from Library RPCI-24

JOURNAL Unpublished  
 COMMENT Other GSSs: RPCI-24-158J11.TV  
 Contact: Shaying Zhao  
 Department of Eukaryotic Genomics  
 The Institute for Genomic Research  
 9712 Medical Center Dr., Rockville, MD 20850, USA  
 Tel: 301 838 0200  
 Fax: 301 838 0208  
 Email: szhao@tigr.org  
 Clones are derived from the mouse BAC library RPCI-24. For BAC library availability, please contact Pieter de Jong (pdejong@mail.cho.org). Clones may be purchased from BACPAC Resources (<http://www.choi.org/bacpac/orderingframe.htm>). BAC end page: [http://www.tigr.org/tdb/bac-ends/mouse/bac\\_end\\_intro.html](http://www.tigr.org/tdb/bac-ends/mouse/bac_end_intro.html)  
 Plate: 158 row: J column: 11  
 Seq primer: SP6  
 Class: BAC ends.

FEATURES  
 source Location/Qualifiers  
 1..713  
 /organism="Mus musculus"  
 /mol\_type="genomic DNA"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="RPCI-24-158J11"  
 /sex="Male"  
 /cell\_type="Spleen/Brain"  
 /clone\_lib="RPCI-24"  
 /note="Vector: pTARBAC1; Site.1: BamHI; Site.2: BamHI; The RPCI-24 mouse BAC library produced by Pieter de Jong. The library was cloned in the pTARBAC1 cloning vector at the BamHI sites using MboI partially digested male C57BL/6J DNA."

BASE COUNT  
 ORIGIN 230 a 154 c 165 g 164 t

Query Match 0.9%: Score 79; DB 28; Length 713;  
 Best Local Similarity 100.0%; Pred. NO. 4.9e-17;  
 Matches 79; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4596 GTGGTGACACAGCGCTTAATCCAGCAGCTTGGAGAGCAGAGCAGATTCTGAGT 4655  
 |||||||  
 DB 396 GTGGTGACACAGCGCTTAATCCAGCAGCTTGGAGAGCAGAGCAGATTCTGAGT 455  
 |||||||

OY 4656 TCAAGGCCAGCTGTGCTA 4674  
 |||||||  
 DB 456 TCAAGGCCAGCTGTGCTA 474  
 |||||||

RESULT 10 671 bp mRNA linear EST 26-OCT-2001  
 BB662563  
 LOCUS BB662563 RIKEN full-length enriched, 15 days embryo head Mus  
 DEFINITION musculus cDNA clone D930008C23 5', mRNA sequence.  
 ACCESSION BB662563  
 VERSION BB662563.1 GI:16496317  
 KEYWORDS EST.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 Mammalia: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi: Mammalia: Eutheria: Rodentia: Sciurognathi: Muridae: Murinae: Mus.  
 1 (bases 1 to 671)  
 AUTHORS Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A.

TITLE  
JOURNAL  
COMMENT

Hiramoto, K., Hori, F., Ishii, Y., Ito, M., Kawai, J., Konno, H., Kouda, M., Koye, S., Matsuyama, T., Miyazaki, A., Nomura, K., Ono, M., Okazaki, Y., Okido, T., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F., Takeda, Y., Tanaka, T., Toya, T., Muramatsu, M. and Hayashizaki, Y.  
RIKEN Mouse ESTs (Arikawa, T., et al. 2001)  
Unpublished

contact: Yoshihide Hayashizaki  
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216  
Email: genome-res@gscc.riken.go.jp,  
URL: http://genome.gsc.riken.go.jp/

Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)

wagi, K., Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Wataniki, M., Tameda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsunura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.

RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)  
Konno, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara, Y. and Hayashizaki, Y.

Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
Kondo, S., Shinagawa, A., Saito, T., Kiyosawa, H., Yamana, I., Aizawa, K., Fukuda, S., Hara, A., Itoh, M., Kawai, J., Shibata, K. and Hayashizaki, Y.

Computational Analysis of Full-length Mouse cDNAs Compared with Human Genome Sequences. Mamm. Genome. 12, 673-677 (2001)  
Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.  
e mouse tissues.

FEATURES  
source

Location/Qualifiers  
1. 671  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="D930008C23"  
/sex="mixed"  
/tissue\_type="head"  
/dev\_stage="15 days embryo"  
/lab\_host="DH10B"  
/clone\_idb="RIKEN full-length enriched, 15 days embryo head"  
/note="Site\_1: Sali; Site\_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5'  
GAGAGAGAGATCTCGAGTAAATTAATTCACCCCCC 3']. cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5'  
GAGAGAGATCTCGAGTAAATTAATTCACCCCCC 3']. cDNA was cloned into the XhoI and BamHI sites. Vector: a modified pBluescript KS(+) after bulk excision from Lambda  
FLC I"

BASE COUNT  
ORIGIN

153 a 147 c 150 g 221 t

Query Match 0.8%; Score 76; DB 10; Length 671;  
Best Local Similarity 100.0%; Pred. No. 5.7e-16;  
Matches 76; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 4638 GCAGGAGATTTCTGAGTTCAGGCCAGCGCTGCTATAGAGTTCAGTCCAGACAGCA 4697  
Db 286 GCGAGGAGATTTCTGAGTTCAGGCCAGCGCTGCTATAGAGTTCAGTCCAGACAGCA 227  
QY 4698 GCGGTACACAGAGAAA 4713  
Db 226 GGGCTACACAGAGAAA 211

RESULT 11  
AK086147/c  
LOCUS 1083 bp mRNA linear HTC 05-DEC-2002  
DEFINITION Mus musculus 15 days embryo head cDNA, RIKEN full-length enriched library, clone:ID930008C23 product:unclassified, full insert sequence.  
ACCESSION AK086147.1 GI:26103252  
VERSION AK086147  
KEYWORDS HTC; CAP trapper.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

REFERENCE  
AUTHORS Carninci, P. and Hayashizaki, Y.  
TITLE High-efficiency full-length cDNA cloning  
JOURNAL Meth. Enzymol. 303, 19-44 (1999)  
MEDLINE 99279253  
PUBMED 10349636

REFERENCE  
AUTHORS Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes  
JOURNAL Genome Res. 10 (10), 1617-1630 (2000)  
MEDLINE 20499374  
PUBMED 11042159

REFERENCE  
AUTHORS Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Katsunagi, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsunoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Wataniki, M., Tameda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsunura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.  
TITLE RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer  
JOURNAL Genome Res. 10 (11), 1757-1771 (2000)  
MEDLINE 20530913  
PUBMED 11076861

REFERENCE  
AUTHORS Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y., Arikawa, T., Hara, A., Fukunishi, Y., Konno, H., Adachi, J., Fukuda, S., Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamana, I., Saito, T., Okazaki, Y., Gojobori, T., Bono, H., Kasukawa, T., Saito, R., Kadoya, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T., Fleischmann, M., Gaasterland, T., Gissi, C., King, B., Kochiwa, H., Kuehl, P., Lewis, S., Matsuo, Y., Nakai, I., Pesole, G., Quackenbush, J., Schirml, L. M., Stabaldi, F., Suzuki, R., Tomita, M., Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H., Balderelli, R., Barsh, G., Blake, J., Boffelli, D., Bojunga, N., Carninci, P., de Bonaldo, M. F., Brownstein, J. D., Bull, C., Fletcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D., Hofmann, M., Hume, D. A., Kamiya, M., Lee, N. H., Lyons, P., Marchionni, L., Mashima, J., Mazzarelli, J., Mombaerts, P., Nordone, P., Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H., Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K. F., Suzuki, H., Toyo-Oka, K., Wang, K. H., Welt, C., Whitlaker, C., Wilmberg, L., Wynshaw-Boris, A., Yoshihide, K., Hasegawa, Y., Kawai, H., Kohlsukl, S.



TITLE and Hayashizaki, Y.  
JOURNAL Functional annotation of a full-length mouse cDNA collection  
MEDLINE Nature 409 (6821), 685-690 (2001)  
PUBMED 21085660  
11217851

REFERENCE 5  
AUTHORS The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.  
TITLE Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs  
JOURNAL Nature 420, 563-573 (2002)  
REFERENCE 6 (bases 1 to 1083)  
AUTHORS Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Horii, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.

COMMENT Direct Submission  
Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@sc.riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)

FEATURES  
source  
1. 1083  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="FANTOM,DB:D930008C23"  
/db\_xref="taxon:10090"  
/clone="D930008C23"  
/tissue\_type="head"  
/clone\_lib="riken full-length enriched mouse cDNA library"  
/dev\_stage="15 days embryo"  
1. 1083  
/note="unclassified"

misc\_feature  
1. 1083  
BASE COUNT 247 a 239 c 241 g 356 t

Query Match 0.88; Score 76; DB 11; Length 1083;  
Best Local Similarity 100.0%; Pred. No. 4,9e-16;  
Matches 76; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 4638 GCAGGCGATTTCGATTCAGGCGAGCTGCTATAGAGTTCAGACACCA 4697  
|||||  
Db 286 GCAGGCGATTTCGATTCAGGCGAGCTGCTATAGAGTTCAGACACCA 227  
|||||

Y 4698 GGGCTACACAGAGAA 4713  
|||||  
Db 226 GGGCTACACAGAGAA 211  
|||||

RESULT 12  
CD538197/c 438 bp mRNA linear EST 11-JUN-2003  
LOCUS B0201B07-3 NIA Mouse Embryonic Germ Cell cDNA Library (long) Mus

ACCESSION musculus cDNA clone NIA:B0201B07 IMAGE:30103506 3', mRNA sequence.  
VERSION CD538197.1 GI:31585932  
KEYWORDS EST.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 438)  
AUTHORS Piao, Y., Ko, N.-T., Lim, M.-K. and Ko, M.-S.-H.  
TITLE Construction of long-transcript enriched cDNA libraries from submicrogram amounts of total RNAs by a universal PCR amplification method  
JOURNAL Genome Res. 11 (9), 1553-1558 (2001)  
MEDLINE 21429098  
PUBMED 11544199  
CONTACT: Dawood B. Dudekula  
COMMENT Laboratory of Genetics  
National Institute on Aging/National Institutes of Health  
333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA  
Email: cdna@igsun.grc.nia.nih.gov  
Plate: B0201 Row: B Column: 07  
Seq primer: -21M13 Forward  
High quality sequence stop: 438  
POLYA-Yes.

FEATURES  
source  
1. 438  
Location/Qualifiers  
1. 438  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="nla:EST:B0201B07-3"  
/db\_xref="taxon:10090"  
/clone="NIA:B0201B07 IMAGE:30103506"  
/sex="male"  
/dev\_stage="embryonic day 8"  
/lab\_host="DH10B"  
/clone\_lib="NIA Mouse Embryonic Germ Cell cDNA Library (long)"  
/note="Vector: PCMV-SPORT6 (Invitrogen); Site 1: SalI; Site 2: NotI; Mouse cDNA project by the Laboratory of Genetics, National Institute on Aging (NIA), Intramural Research Program, NIH (http://igsun.grc.nia.nih.gov/cDNA). This is a long-transcript enriched cDNA library (Ref. Genome Res. 11: 1553-1558 (2001)). (PMID: 11544199). Total RNAs were obtained from Dr. Mark G. Carter (NIH/NIA-IRP). EG cells were cultured at 37 C, 5% CO2 in DMEM supplemented with 15% ES cell qualified FBS, 0.1mM non-essential amino acids, 2 mM glutamine, penicillin/streptomycin, 1 mM sodium pyruvate, 0.1 mM beta-mercaptoethanol, and 10000000 units of LIF per liter. Double-stranded cDNAs were synthesized with an Oligo(dT) primer [Invitrogen]: 5'-pGACTAGTCTAGATCCGAGCGCGCCCTTTTTTT-3' from 2.5 ug of total RNA, treated with T4 DNA polymerase, and purified by ethanol-precipitation. The cDNAs were ligated to lone-linker Lf-SalI, purified by phenol/chloroform, and separated from free linkers by Centricon 100. Then, the cDNAs were amplified by long-range high fidelity PCR using Ex Taq polymerase (Takara) with a primer SalI-S. The products were purified by phenol/chloroform and Centricon 100. The cDNAs were digested with SalI and NotI enzymes and cloned into SalI/NotI site of PCMV-SPORT6 plasmid vector. The DH10B E. coli host was transformed with the ligation mixture by the standard chemical method. The average insert size is about 4.0 kb. The library was constructed by Yulan Piao."

BASE COUNT 118 a 88 c 100 g 132 t

Query Match 0.88; Score 75; DB 14; Length 438;  
Best Local Similarity 100.0%; Pred. No. 1.5e-15;  
Matches 75; Conservative 0; Mismatches 0; Indels 0; Gaps 0;



OY		4645	GATTTCCTAGTCAAGGCGACCCGTGCCTATAGAAGTGAGTTCCAGGACGCCAGGGCGTAC	4704
Db		79	GATTTCCTAGTCAAGGCGACCCGTGCCTATAGAAGTGAGTTCCAGGACGCCAGGGCGTAC	20
OY		4705	ACAGAGAAACCCTGT	4719
Db		19	ACAGAGAAACCCTGT	5
RESULT 13				
AZ427733				
DEFINITION			540 bp DNA linear GSS 03-OCT-2000	
LOCUS			1M209H24F Mouse 10kb plasmid U08C1M library Mus musculus genomic	
ACCESSION			clone U08C1M209H24 R, genomic survey sequence.	
VERSION			A2427733.1	
KEYWORDS			GI:10551746	
SOURCE			GSS.	
ORGANISM			Mus musculus (house mouse)	
			Mus musculus	
			Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
			Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.	
REFERENCE			I (bases 1 to 540)	
AUTHORS			Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamll,C.,	
			Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly	
			, M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.	
			and Wright,D., Weiss,R.	
TITLE			Mouse whole genome scaffolding with paired end reads from 10xb	
JOURNAL			plasmid inserts	
COMMENT			unpublished	
			Contact: Robert B. Weiss	
			University of Utah Genome Center	
			University of Utah	
			Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT	
			84112, USA	
			Tel: 801 585 5606	
			Fax: 801 585 7177	
			Email: ddunn@genetics.utah.edu	
			Insert Length: 10000 Std Error: 0.00	
			Plate: 0209 row: H column: 24	
			Seq primer: CACACAGGAACAACGACTCATGACC	
			Class: plasmid ends	
FEATURES			High quality sequence stop: 540.	
source			Location/Qualifiers	
			1..540	
			/organism="Mus musculus"	
			/mol_type="genomic DNA"	
			/strain="C57BL/6J"	
			/db_xref="taxon:10090"	
			/clone="U08C1M209H24"	
			/sex="Male"	
			/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"	
			/clone_lib="Mouse 10kb plasmid U08C1M library"	
			/note="Vector: pMD42nv; Purified genomic DNA from M.	
			musculus C57BL/6J (male) was obtained from the Jackson	
			Laboratory Mouse DNA Resource	
			(http://www.jax.org/resources/documents/dnares/). The DNA	
			was hydrodynamically sheared by repeated passage through a	
			0.005 inch orifice at constant velocity. The sheared DNA	
			was blunt end-repaired with T4 DNA polymerase and T4	
			polynucleotide kinase. Adaptor oligonucleotides were	
			ligated to the blunt ends in high molar excess. The	
			adapted DNA was purified and size-selected for a 9.5 to	
			10.5 kb range using preparative agarose gel	
			electrophoresis. Vector DNA was prepared from a derivative	
			of pMD42 (gII4732114(gybIARf29072.1)), a copy-number	
			inducible derivative of plasmid RI. The vector was ligated	
			with adaptors complementary to the insert adaptors and	
			purified. The sheared, adapted mouse DNA was annealed to	
			adapted vector DNA, and transformed into	
			chemically-competent E. coli XL10-Gold (Stratagene) cells	
			and selected for ampicillin resistance."	
BASE COUNT			188 a 113 c 122 g 116 t	
ORIGIN			1 others	

Qy	4645	GATTTCTGAGTTCAGGCGCAGCTGGTCTATAGAGTTCAGAGCAGCGGCTAC	4704
Db	65		
Qy	4705	ACAGAGAAACCCCTGT	4719
Db	125		
Db	125	ACAGAGAAACCCCTGT	139
RESULT 14			
CNS02JZY			
LOCUS			
DEFINITION	CNS02JZY	674 bp	DNA
LOCATION	Tetradon nigroviridis genome survey sequence T7 end of clone		
FEATURES	144F01 of library G from Tetradon nigroviridis, genomic survey		
ACCESSION	AL200887.1	GI:7859232	
VERSION			
KEYWORDS	GSS; genome survey; sequence.		
SOURCE	Tetradon nigroviridis		
ORGANISM	Tetradon nigroviridis		
REFERENCE	Euhariyola, Metazoa; Chordata; Vertebrata; Euteleostomi; Acanthopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes; Tetraodontoidae; Tetraodon.		
AUTHORS	1 Roest Crolius, H., Jallion, O., Dasilva, C., Bouneau, L., Fisher, C., Bernot, A., Fitzames, C., Wincker, P., Brothier, P., Quetier, F., Saurin, W. and Weissenbach, J.		
TITLE	Estimate of human gene number provided by genome-wide analysis using Tetradon nigroviridis DNA sequence		
JOURNAL	Nat. Genet. 25 (2), 235-238 (2000)		
MEDLINE	20296633		
PUBMED	10835645		
REFERENCE	2		
AUTHORS	Roest Crolius, H., Jallion, O., Dasilva, C., Ozouf-Costaz, C., Fitzames, C., Fischer, C., Bouneau, L., Billault, A., Quetier, F., Saurin, W., Bernot, A. and Weissenbach, J.		
TITLE	Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetradon nigroviridis		
JOURNAL	Genome Res. 10 (7), 939-949 (2000)		
MEDLINE	20359837		
PUBMED	10899143		
REFERENCE	3 (bases 1 to 674)		
AUTHORS	Genoscope.		
TITLE	Submitted (12-APR-2000) Genoscope - Centre National de Sequence :		
JOURNAL	BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr		
COMMENT	- Web : www.genoscope.cns.fr)		
FEATURES	This sequence is a single read and was generated as part of a large scale clone-and-sequencing project of the Tetradon nigroviridis genome. For more information, please take a look at		
SOURCE	http://www.genoscope.cns.fr/Tetradon.		
location/Qualifiers			
1. 674			
/organism="Tetradon nigroviridis"			
/mol_type="genomic DNA"			
/db_xref="taxon:99883"			
/clone="144F01"			
/clone_lib="G"			
/note="Genoscope sequence ID : C0KGI44CC01LP1-end : 77"			
BASE COUNT	121 a	127 c	219 g
ORIGIN	180 t	27 others	
Query Match	0.8%;	Score 75;	DB 28; Length 540;
Best Local Similarity	100.0%;	Prod. No. 1,4e-15;	
Matches : 75;	Conservative 0;	Mismatches 0;	Indels 0;
Gaps 0;			



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Comphen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 13, 2003, 12:10:00 ; Search time 525 seconds  
(without alignments)  
7562.364 Million cell updates/sec

Title: US-09-877-935-1

Perfect score: 8995  
Sequence: 1 gatctggtgaccacgaagaca.....cttaagctctgcacacatg 8995

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 569978 seqs, 220691566 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

Issued\_Patents\_NA:\*  
1: /cgn2\_6/ptodata/1/lna/5A\_COMB.seq:\*  
2: /cgn2\_6/ptodata/1/lna/5B\_COMB.seq:\*  
3: /cgn2\_6/ptodata/1/lna/6A\_COMB.seq:\*  
4: /cgn2\_6/ptodata/1/lna/6B\_COMB.seq:\*  
5: /cgn2\_6/ptodata/1/lna/PCTUS\_COMB.seq:\*  
6: /cgn2\_6/ptodata/1/lna/backfile1.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	64	0.7	3588	1	US-08-197-792-32
C 2	64	0.7	3588	1	US-08-459-850-32
C 3	64	0.7	3588	1	US-08-459-214-32
C 4	63	0.7	80246	3	US-09-078-294-4
C 5	62	0.7	3381	3	US-09-009-119-1
C 6	62	0.7	3381	4	US-09-371-507-1
C 7	62	0.7	3383	5	PCT-US95-09098-1
C 8	61	0.7	338	4	US-09-495-050A-105
C 9	61	0.7	818	3	US-08-205-697A-4
C 10	61	0.7	818	3	US-08-702-525-4
C 11	61	0.7	818	5	PCT-US95-02576-4
C 12	61	0.7	1570	3	US-08-205-697A-10
C 13	61	0.7	1570	3	US-08-702-525-10
C 14	61	0.7	1570	3	PCT-US95-02576-10
C 15	61	0.7	1606	3	US-08-702-525-64
C 16	61	0.7	1606	5	PCT-US95-02576-64
C 17	61	0.7	1816	4	US-09-645-926A-5
C 18	61	0.7	1888	3	US-08-702-697A-1
C 19	61	0.7	1888	3	US-08-702-525-1
C 20	61	0.7	1888	5	PCT-US95-02576-1
C 21	61	0.7	2516	3	US-08-205-697A-3
C 22	61	0.7	2516	3	US-08-702-525-3
C 23	61	0.7	2516	5	PCT-US95-02576-3
C 24	61	0.7	2529	3	US-09-051-969A-5
C 25	61	0.7	4157	2	US-08-871-266B-1
C 26	61	0.7	4157	2	US-08-819-458A-1
C 27	61	0.7	4157	2	US-09-018-864A-1

C 28	61	0.7	4157	3	US-08-871-267B-1	Sequence 1, Appl
C 29	61	0.7	4157	3	US-09-618-419-1	Sequence 1, Appl
C 30	61	0.7	55298	4	US-09-491-356C-1	Sequence 1, Appl
C 31	61	0.7	80595	3	US-09-078-294-3	Sequence 3, Appl
C 32	59	0.7	223	1	US-08-222-177A-14	Sequence 14, Appl
C 33	59	0.7	1920	4	US-09-534-638-6	Sequence 6, Appl
C 34	59	0.7	9840	4	US-09-534-638-1	Sequence 1, Appl
C 35	57	0.6	57	1	US-08-222-177A-128	Sequence 128, Appl
C 36	57	0.6	113	4	US-09-354-147C-36	Sequence 36, Appl
C 37	56	0.6	747	4	US-09-220-132-92	Sequence 92, Appl
C 38	56	0.6	2775	1	US-08-149-096A-1	Sequence 1, Appl
C 39	56	0.6	63588	4	US-09-873-404-3	Sequence 3, Appl
C 40	55	0.6	211	4	US-09-354-147C-29	Sequence 29, Appl
C 41	55	0.6	10409	3	US-08-772-440-33	Sequence 33, Appl
C 42	55	0.6	176373	3	US-09-128-155-17	Sequence 17, Appl
C 43	54	0.6	2121	4	US-09-425-488-1	Sequence 1, Appl
C 44	54	0.6	2235	3	US-09-334-601-3	Sequence 3, Appl
C 45	54	0.6	2360	3	US-08-916-043-3	Sequence 3, Appl

#### ALIGNMENTS

RESULT 1  
US-08-197-792-32/C  
Sequence 32, Application US/08197792  
Patent No. 5525488  
GENERAL INFORMATION:  
APPLICANT: Anthony J. Mason  
APPLICANT: Peter H. Seeburg  
TITLE OF INVENTION: Nucleic Acid Encoding the Alpha or Beta Chains of Inhibin a  
NUMBER OF SEQUENCES: 44  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genentech, Inc.  
STREET: 460 Point San Bruno Blvd  
CITY: South San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94080  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: patin (Genentech)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/197,792  
FILING DATE: 16-FEB-1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/958414  
FILING DATE: 08-OCT-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/744207  
FILING DATE: 12-AUG-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/215466  
FILING DATE: 05-JUL-1988  
APPLICATION NUMBER: 07/906729  
FILING DATE: 31-DEC-1986  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 06/827710  
FILING DATE: 07-FEB-1986  
APPLICATION NUMBER: 06/783910  
FILING DATE: 03-OCT-1985  
AUTORNEY/AGENT INFORMATION:  
NAME: Hasak, Janet E.  
REGISTRATION NUMBER: 28,616  
REFERENCE/DOCKET NUMBER: 297P2D4  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415/225-1896  
TELEFAX: 415/952-9881





```

1 STREET: 8110 Gatehouse Road, Suite 500 East
2 CITY: Falls Church
3 STATE: Virginia
4 COUNTRY: U.S.A.
5 ZIP: 22042
6
7 COMPUTER READABLE FORM:
8 MEDIUM TYPE: Floppy disk
9 COMPUTER: IBM PC compatible
10 OPERATING SYSTEM: PC-DOS/MS-DOS
11 SOFTWARE: PatentIn Release #1.0, Version #1.30
12
13 CURRENT APPLICATION DATA:
14 APPLICATION NUMBER: PCT/US95/09098
15 FILING DATE: 20-JUL-1995
16 CLASSIFICATION:
17
18 ATTORNEY/AGENT INFORMATION:
19 NAME: Murphy Jr., Gerald M.
20 REGISTRATION NUMBER: 28,977
21 REFERENCE/DOCKET NUMBER: 2185-110P
22 TELECOMMUNICATION INFORMATION:
23 TELEPHONE: (703) 205-8000
24 TELEFAX: (703) 205-8050
25 TELEX: 248345
26
27 INFORMATION FOR SEQ ID NO: 1:
28
29 SEQUENCE CHARACTERISTICS:
30 LENGTH: 3383 base pairs
31 TYPE: nucleic acid
32 STRANDEDNESS: double
33 TOPOLOGY: linear
34 MOLECULE TYPE: DNA (genomic)
35
36 HYPOTHETICAL: NO
37
38 ANTI-SENSE: NO
39
40 ORIGINAL SOURCE:
41 ORGANISM: Chlamydomonas reinhardtii
42 STRAIN: RS-3
43
44 PCT-US95-09098-1
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
9
```

[illegible]











GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 13, 2003, 12:10:00 ; Search time 2039 Seconds  
(without alignments)  
11908.507 Million cell updates/sec

Title: US-09-877-935-1

Perfect score: 8995  
Sequence: 1 gatctgtgtcaccacgaagaca.....ctctagctctcaccatg 8995

Scoring table: OLIGO\_NDC  
Gapop 60.0 , Gapext 60.0

Searched: 2552756 segs, 1349719017 residues

Word size : 0

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : N\_Geneseq\_19Jun03:\*

1: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1980.DAT:\*  
2: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1981.DAT:\*  
3: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1982.DAT:\*  
4: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1983.DAT:\*  
5: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1984.DAT:\*  
6: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1985.DAT:\*  
7: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1986.DAT:\*  
8: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1987.DAT:\*  
9: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1988.DAT:\*  
10: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1989.DAT:\*  
11: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1990.DAT:\*  
12: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1991.DAT:\*  
13: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1992.DAT:\*  
14: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1993.DAT:\*  
15: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1994.DAT:\*  
16: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1995.DAT:\*  
17: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1996.DAT:\*  
18: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1997.DAT:\*  
19: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1998.DAT:\*  
20: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1999.DAT:\*  
21: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2000.DAT:\*  
22: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:\*  
23: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT:\*  
24: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2002.DAT:\*  
25: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2003.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	8995	100.0	8995	21	AAA48515 Murine villin gene
2	8631	96.2	8993	21	AAA49605 Mouse villin gene
3	66	0.7	2628	24	AA519044 Mouse WU-1 haemato
4	65	0.7	1623	24	AB565621 Mouse Fabp gene (E
5	64	0.7	574	21	ABN81091 Shrimp polynucleot
6	64	0.7	2447	22	AA161087 Human polynucleoti
7	63	0.7	80	23	AA513715 Simple sequence re
8	63	0.7	279	21	AAA74225 Lobliolly pine SSR

9	63	0.7	382	15	AA055246 Grapevine ribosoma
10	63	0.7	382	21	AAA74316 Lobliolly pine SSR
11	63	0.7	414	22	AA193748 Human polynucleoti
12	63	0.7	4580	17	AA132034 Proliferation-inhi
13	63	0.7	80240	20	AA083940 NC-ctrlg derived
14	62	0.7	219	13	AA027094 Fragile X diagnost
15	62	0.7	381	24	ABX04909 Conus sp conotoxin
16	62	0.7	463	21	ABN81124 Shrimp polynucleot
17	62	0.7	2573	19	AAVA6297 C. reinhardtii pro
18	62	0.7	3316	23	AB110742 Drosophila melanog
19	62	0.7	3381	18	AA178597 Porphyritin accumula
20	62	0.7	3381	19	AAVA6298 C. reinhardtii pro
21	62	0.7	3383	18	AA160965 Porphyritin-accumula
22	62	0.7	14704	13	AA020685 PKS 741 insert con
23	62	0.7	14704	13	AA020685 PKS 741 insert con
24	62	0.7	42999	24	AB565032 Invertebrate forag
25	62	0.7	173810	24	ABN85752 Mouse chromosome 1
26	61	0.7	66	23	AA513714 Simple sequence re
27	61	0.7	68	21	AA298506 H. discus derived
28	61	0.7	76	23	AA513704 Simple sequence re
29	61	0.7	82	21	AA298483 H. discus derived
30	61	0.7	90	23	AA513709 Simple sequence re
31	61	0.7	127	25	AB219796 Group III CDNA can
32	61	0.7	140	22	AAK90823 Human digestive sy
33	61	0.7	219	21	ABN81161 Shrimp polynucleot
34	61	0.7	266	13	AA027097 Fragile X diagnost
35	61	0.7	282	13	AA027100 Fragile X diagnost
36	61	0.7	36	22	AAH29391 Drosophila melanog
37	61	0.7	362	21	AAA74223 Lobliolly pine SSR
38	61	0.7	407	22	AAE66146 Novel human polynu
39	61	0.7	448	18	AAE68682 Peneaus vannamei g
40	61	0.7	475	22	AB443925 Human breast cell
41	61	0.7	475	22	AB443925 Human breast cell
42	61	0.7	475	22	AB443925 Human fetal liver
43	61	0.7	475	22	AB443925 Human fetal liver
44	61	0.7	475	22	AAK02673 Probe #2623 for ge
45	61	0.7	475	22	AAK28111 Human bone marrow
					Probe #2621 for ge

ALIGNMENTS

RESULT 1	
AAA48515	AAA48515 standard; DNA; 8995 BP.
XX	XX
AC	AAA48515;
XX	XX
DT	19-DEC-2000 (first entry)
XX	XX
DE	Murine villin gene regulatory region.
XX	XX
KW	Mouse; villin; regulatory region; digestive tract;
KW	colorectal cancer mouse model; ds.
XX	XX
OS	Mus sp.
XX	XX
FH	FH
FT	Key
FT	misc-signal
FT	Location/Qualifiers
FT	3442
FT	/tag= a
FT	/note= "transcriptional start site"
FT	3442..3487
FT	/tag= b
FT	/number= 1
FT	3488..8981
FT	/tag= c
FT	/number= 1
FT	8982..8995
FT	/tag= d
FT	/number= 2
FT	8993
FT	/tag= e
FT	/note= "translational start site"



OY	1681	TTGACATGTCGTGTAGAAAGATAGAAAAATGAAGACTTGAAGTCTAGTCTCTGAAACCCACA	1740
Db	1681	TTAGCATGTGCTGTGAGAAAGATAGAAAAATGAAGACTTGAAGTCTAGTCTCTGAAACCCACA	1740
OY	1741	GAGGAGGCGGAAGAACCCCACTGTAAGTGTGCTGTGAGGCTTACATACAACTTCACAT	1800
Db	1741	GAGGAGGCGGAAGAACCCCACTGTAAGTGTGCTGTGAGGCTTACATACAACTTCACAT	1800
OY	1801	AATAGTTCATATGATTAATTAATTAATTTTAAAGGATATATGTGGAGGGA	1860
Db	1801	AATAGTTCATATGATTAATTAATTAATTTTAAAGGATATATGTGGAGGGA	1860
OY	1861	GAGATGCTCAGCTTCGAGAGCACTTGCTGCTTGTGAGAGGACTTAGATTCAGTTCCC	1920
Db	1861	GAGATGCTCAGCTTCGAGAGCACTTGCTGCTTGTGAGAGGACTTAGATTCAGTTCCC	1920
OY	1921	AGGACTCATATGAGTGGGCTCACAGCCATGCTTAATCCAGTTCGAGAGGCTTCCACACCT	1980
Db	1921	AGGACTCATATGAGTGGGCTCACAGCCATGCTTAATCCAGTTCGAGAGGCTTCCACACCT	1980
OY	1981	CTTCTGGCCTCCACAGGACACACATACATAGTACACAGACATACATGACAGCAAAACAC	2040
Db	1981	CTTCTGGCCTCCACAGGACACACATACATAGTACACAGACATACATGACAGCAAAACAC	2040
OY	2041	CATACACACATTAATTAATTAAGAAACTTAAAGGTCATGTGTGTTGTTAAACATTTGCT	2100
Db	2041	CATACACACATTAATTAATTAAGAAACTTAAAGGTCATGTGTGTTGTTAAACATTTGCT	2100
OY	2101	TACACATGCTATTTGAACATGTACAAACGACACACACTGAAGAGGATCTGGGGTGGAG	2160
Db	2101	TACACATGCTATTTGAACATGTACAAACGACACACACTGAAGAGGATCTGGGGTGGAG	2160
OY	2161	AGATGGCTCAGCGGTTTAAAGACACATGACTGCTCTCCGAAAGAGGTCCTAGTTCAAT	2220
Db	2161	AGATGGCTCAGCGGTTTAAAGACACATGACTGCTCTCCGAAAGAGGTCCTAGTTCAAT	2220
OY	2221	CTTAGCAACACATAGTGGGCTCACACCATCATATAGATCTGACACCCCTCTTGCT	2280
Db	2221	CTTAGCAACACATAGTGGGCTCACACCATCATATAGATCTGACACCCCTCTTGCT	2280
OY	2281	GCATGTGAAGCAGGTGAGAGCTACAGTGTACTAGATATCTAATAATTAATCTTT	2340
Db	2281	GCATGTGAAGCAGGTGAGAGCTACAGTGTACTAGATATCTAATAATTAATCTTT	2340
OY	2341	TTTAAAAAATGAAGAGGATCTGAGACACCTCAAAAGAGATTTATGACAGTACTCAG	2400
Db	2341	TTTAAAAAATGAAGAGGATCTGAGACACCTCAAAAGAGATTTATGACAGTACTCAG	2400
OY	2401	GGTGATTTATCTATCCTGAGAGTTTTTCCCTTCCGCTTGCCACTGGGTGGAGACAG	2460
Db	2401	GGTGATTTATCTATCCTGAGAGTTTTTCCCTTCCGCTTGCCACTGGGTGGAGACAG	2460
OY	2461	CCCCCTTTCATTCACAAAGAGGGGTCTACATATTTTGTGAACAAACAGCACCTGGCAGT	2520
Db	2461	CCCCCTTTCATTCACAAAGAGGGGTCTACATATTTTGTGAACAAACAGCACCTGGCAGT	2520
OY	2521	ATGTTTACTGTCCTTGTGACTATGAGCAGCGCAGCGCGCGCGCGCACACACACAC	2580
Db	2521	ATGTTTACTGTCCTTGTGACTATGAGCAGCGCAGCGCGCGCGCGCGCACACACACAC	2580
OY	2581	ACACACACACACACACACACACACACACATTCAGTCTCCAGAGCTCTTGGGAAGTCA	2640
Db	2581	ACACACACACACACACACACACACACACATTCAGTCTCCAGAGCTCTTGGGAAGTCA	2640
OY	2641	AGAGAGGCTGCCCTCAAAACAGATCTTCATCTTCCCTCTTAAAGAGACACAGATTC	2700
Db	2641	AGAGAGGCTGCCCTCAAAACAGATCTTCATCTTCCCTCTTAAAGAGACACAGATTC	2700
OY	2701	AAGGTGCGAAGATCTACAGGGGGCAGAGGCGAGGGAGGGAGGCCATGGTTTTCC	2760
Db	2701	AAGGTGCGAAGATCTACAGGGGGCAGAGGCGAGGGAGGGAGGCCATGGTTTTCC	2760

```
Db 3841 TCACTGTACACATCTGAATGTTCTTGTGTGCACATTCCTCCCTACACATCTCTGGGA 3900
OY 3901 GGTCTATCTCTGGACATGTATCTCTGGATGTAACTGCAGCCACCGAGAGAGAGGG 3960
Db 3901 GGTCTATCTCTGGACATGTATCTCTGGATGTAACTGCAGCCACCGAGAGAGAGGG 3960
OY 3961 AGAGTCAGAGAGCTGTCTTCCCTAGGCGCTATTAGGCGTGCACATCACCCCTTCTAGAAAT 4020
Db 3961 AGAGTCAGAGAGCTGTCTTCCCTAGGCGCTATTAGGCGTGCACATCACCCCTTCTAGAAAT 4020
OY 4021 GGGCCCTCCATTTTTCGGTTACCATGATCTATTATTATACAGAGTGGGCGAGTGAAGCCA 4080
Db 4021 GGGCCCTCCATTTTTCGGTTACCATGATCTATTATTATACAGAGTGGGCGAGTGAAGCCA 4080
OY 4081 AACCTGCCAGAAAGTTGGAGACTCAGTGCAGACCAAGTTATCTGCTCAGAAAATCCCGCTG 4140
Db 4081 AACCTGCCAGAAAGTTGGAGACTCAGTGCAGACCAAGTTATCTGCTCAGAAAATCCCGCTG 4140
OY 4141 TCACCTTAGAGTTGGAGAAATCTGCTTGGGGGCTTCCAGAGTCTTGGTTAGCAGAGAGGT 4200
Db 4141 TCACCTTAGAGTTGGAGAAATCTGCTTGGGGGCTTCCAGAGTCTTGGTTAGCAGAGAGGT 4200
OY 4201 ATCCCTTGTATAGGCGATGACCTAGTCTATGCTTACTACATTCCTGTCCAGTTAAAG 4260
Db 4201 ATCCCTTGTATAGGCGATGACCTAGTCTATGCTTACTACATTCCTGTCCAGTTAAAG 4260
OY 4261 CTGGAACATAAAACCCAGCGGACGCGCCAGAGATTCTCTACAGTTGTACCCCAAGAACACA 4320
Db 4261 CTGGAACATAAAACCCAGCGGACGCGCCAGAGATTCTCTACAGTTGTACCCCAAGAACACA 4320
OY 4321 AGACAGTATGATATGCAAGGATATGATGCTGGGAGAGAGAGAACTTAAACCCCAAG 4380
Db 4321 AGACAGTATGATATGCAAGGATATGATGCTGGGAGAGAGAGAACTTAAACCCCAAG 4380
OY 4381 GCGCCAGAGTCCGTTCCCTAGTTGCAATGCGACATGAGTGTAGTCTAGTATAGGCGTG 4440
Db 4381 GCGCCAGAGTCCGTTCCCTAGTTGCAATGCGACATGAGTGTAGTCTAGTATAGGCGTG 4440
OY 4441 TGAGTTGGTAGTACACAGCATGATGATGTTTCATGCTGTAGTGTATTAATCTAGACAC 4500
Db 4441 TGAGTTGGTAGTACACAGCATGATGATGTTTCATGCTGTAGTGTATTAATCTAGACAC 4500
OY 4501 TTGGAGAGCTGACAGCAGAGAGATTGCTATATGTTTGAGGCCAGCCTGAGCTATAGACGA 4560
Db 4501 TTGGAGAGCTGACAGCAGAGAGATTGCTATATGTTTGAGGCCAGCCTGAGCTATAGACGA 4560
OY 4561 GACTTTTCTTTAAGAAAAAATGAAGCCAGCAGCTGTGSCAACCCCTTATATCCCA 4620
Db 4561 GACTTTTCTTTAAGAAAAAATGAAGCCAGCAGCTGTGSCAACCCCTTATATCCCA 4620
OY 4621 GCACCTTGGAGGACAGAACAGCAGCATTTCTGAGTTCAAGGCCACCTGGTCTATAGAGT 4680
Db 4621 GCACCTTGGAGGACAGAACAGCAGCATTTCTGAGTTCAAGGCCACCTGGTCTATAGAGT 4680
OY 4681 GAGTTCCAGGACAGCAGGCGCTACACAGAGAAACCCTGTTTGAAGAAAAACAGAAAAACA 4740
Db 4681 GAGTTCCAGGACAGCAGGCGCTACACAGAGAAACCCTGTTTGAAGAAAAACAGAAAAACA 4740
OY 4741 AACAAAAAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 4800
Db 4741 AACAAAAAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 4800
OY 4801 TAGGCTGTCTGTCTAGTGTAGTGTAGTGTGGGAGCTTCAAGACTTATATTAATAGGCC 4860
Db 4801 TAGGCTGTCTGTCTAGTGTAGTGTAGTGTGGGAGCTTCAAGACTTATATTAATAGGCC 4860
OY 4861 TTTTATATCTGTGTACAGACAGAGAAAGTTTCACTGTGGGACACAGTGGGACCTTGAGA 4920
Db 4861 TTTTATATCTGTGTACAGACAGAGAAAGTTTCACTGTGGGACACAGTGGGACCTTGAGA 4920
OY 4921 AAGTACTCTTGTCCAGCCCAAAAATTTCTGGGAAGGCTTCTGGAGAGTGTGTCCCAT 4980
Db 4921 AAGTACTCTTGTCCAGCCCAAAAATTTCTGGGAAGGCTTCTGGAGAGTGTGTCCCAT 4980
```

```
Db 4921 AAGTACTCTTGTCCAGCCCAAAAATTTCTGGGAAGGCTTCTGGAGAGTGTGTCCCAT 4980
OY 4981 CAGACTACTGTCTCTGAAGGCGAGAGAGAGGTTGGAAGAAATGTTGTGACAGACAGTT 5040
Db 4981 CAGACTACTGTCTCTGAAGGCGAGAGAGAGGTTGGAAGAAATGTTGTGACAGACAGTT 5040
OY 5041 GGAACAGAAAGCAGAGAGGAGGAGCATCCAGATTTCTGAACATGTAGCTGACTTGTGT 5100
Db 5041 GGAACAGAAAGCAGAGAGGAGGAGCATCCAGATTTCTGAACATGTAGCTGACTTGTGT 5100
OY 5101 TCTCTGGGTGACAAATGTTCCTCCAGGATAGGGGCTGTAGAAAGGGGACAGGGGTGACC 5160
Db 5101 TCTCTGGGTGACAAATGTTCCTCCAGGATAGGGGCTGTAGAAAGGGGACAGGGGTGACC 5160
OY 5161 AATGAGTTCAAGTTAGGAGACATCCAGCCAGGCTCTTGTCTGCAAGCTAAAGATG 5220
Db 5161 AATGAGTTCAAGTTAGGAGACATCCAGCCAGGCTCTTGTCTGCAAGCTAAAGATG 5220
OY 5221 AGAGCCCTCTAAACCTCCCTGAAGTTTAGGGAGACAGAGAGCTGAGAGAGATCCTTCTA 5280
Db 5221 AGAGCCCTCTAAACCTCCCTGAAGTTTAGGGAGACAGAGAGCTGAGAGAGATCCTTCTA 5280
OY 5281 GGGTGAAGAGAGGATATGCTGCTGACCAACATGCTAGAGGACAAAGCAGTTGGACAG 5340
Db 5281 GGGTGAAGAGAGGATATGCTGCTGACCAACATGCTAGAGGACAAAGCAGTTGGACAG 5340
OY 5341 TTACCCCTCAGAACCGACCATCCCTCTTGGCTTAGAGAGGCTGGGCGCTTCTGTTT 5400
Db 5341 TTACCCCTCAGAACCGACCATCCCTCTTGGCTTAGAGAGGCTGGGCGCTTCTGTTT 5400
OY 5401 AAGAATCTTACTTTTCTCAGAGAGAGCAGCAGACCTTTTGTCCCTCTCTGTGTCAA 5460
Db 5401 AAGAATCTTACTTTTCTCAGAGAGAGCAGCAGACCTTTTGTCCCTCTCTGTGTCAA 5460
OY 5461 TAAACACCCCTGTGTGAACATATTATTATTAGTGCAGTTGCTCAGAGACGTCCA 5520
Db 5461 TAAACACCCCTGTGTGAACATATTATTATTAGTGCAGTTGCTCAGAGACGTCCA 5520
OY 5521 TCTGTAGACCTCTGCTCTTAACCTCACCAGGATATGGCCACATTCCTCACCCAGAAAG 5580
Db 5521 TCTGTAGACCTCTGCTCTTAACCTCACCAGGATATGGCCACATTCCTCACCCAGAAAG 5580
OY 5581 TGCAGAGAGAGCCTTACAGAAAGGCTAAACAGTAAACAAAGATGCCAGATTAACAA 5640
Db 5581 TGCAGAGAGAGCCTTACAGAAAGGCTAAACAGTAAACAAAGATGCCAGATTAACAA 5640
OY 5641 ACTATATCTTTGTAACCAATTTGCTTGTGTAACAGAGAGGCGGTGTGAGTGTAT 5700
Db 5641 ACTATATCTTTGTAACCAATTTGCTTGTGTAACAGAGAGGCGGTGTGAGTGTAT 5700
OY 5701 GTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 5760
Db 5701 GTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 5760
OY 5761 CTTGGGGACCTTTTATGCTTAAAGAAATATCTGATTTTGGCGCCATGCAACAGGGGTAT 5820
Db 5761 CTTGGGGACCTTTTATGCTTAAAGAAATATCTGATTTTGGCGCCATGCAACAGGGGTAT 5820
OY 5821 TGGGAGAGTCAAGGCTTCTGCAACACAGTAAAGTGGCCAGAGATGGTTGGGCTGAA 5880
Db 5821 TGGGAGAGTCAAGGCTTCTGCAACACAGTAAAGTGGCCAGAGATGGTTGGGCTGAA 5880
OY 5881 TCACCAAGGGGAGGCTGATCAGATGAGACAGAAATCACAAGATTAAGCCACCTGTGG 5940
Db 5881 TCACCAAGGGGAGGCTGATCAGATGAGACAGAAATCACAAGATTAAGCCACCTGTGG 5940
OY 5941 GCTCAGAAAGAGGAGTTTACAAAGAGGTAAAGGCCAAGCATTATATCCAAAGATAGC 6000
Db 5941 GCTCAGAAAGAGGAGTTTACAAAGAGGTAAAGGCCAAGCATTATATCCAAAGATAGC 6000
OY 6001 TCAAAATCAAAAGTGAAGAGAGATTAAGTGAAGAGATGGGGCTCAGTGTGGGACACC 6060
Db 6001 TCAAAATCAAAAGTGAAGAGAGATTAAGTGAAGAGATGGGGCTCAGTGTGGGACACC 6060
```



OY 6061 TGACCTTGCATTATTAGTCACTAGAGCCAAAGAGCAGTCACAGAGGGTGAAGTGGCTCTA 6120  
|||||  
Db 6061 TGACCTTGCACCTATTAGTCACTAGAGCCAAAGAGCAGTCACAGAGGGTGAAGTGGCTCTA 6120  
OY 6121 CTCAGCTTGGAGCAGCAGCAGTGGAGAAATGGGTGACTCCATCTGTATGAGAGGGCTGAG 6180  
|||||  
Db 6121 CTCAGCTTGGAGCAGCAGCAGTGGAGAAATGGGTGACTCCATCTGTATGAGAGGGCTGAG 6180  
OY 6181 CACACACAGGTACAAAGTGTCCCTGTGTCTCATGACAGATTCCTGGCCAGTTTCAAG 6240  
|||||  
Db 6181 CACACACAGGTACAAAGTGTCCCTGTGTCTCATGACAGATTCCTGGCCAGTTTCAAG 6240  
OY 6241 GACTAAGAGATCATCTCTGTGGGAAACAAAGTATCCAAAGCCCTTAAGCCCATTTGGTCT 6300  
|||||  
Db 6241 GACTAAGAGATCATCTCTGTGGGAAACAAAGTATCCAAAGCCCTTAAGCCCATTTGGTCT 6300  
OY 6301 AATTAAATCAGAAACCCCTGGGAGTCAGAGCTGTGACAGCAGAGAGCTTTTAAAAAGCTC 6360  
|||||  
Db 6301 AATTAAATCAGAAACCCCTGGGAGTCAGAGCTGTGACAGCAGAGAGCTTTTAAAAAGCTC 6360  
OY 6361 CCAGGTGATTTCGATCAGAGCTGGAACAAACACAGCTACAGSTTGAACAGAGAGGC 6420  
|||||  
Db 6361 CCAGGTGATTTCGATCAGAGCTGGAACAAACACAGCTACAGSTTGAACAGAGAGGC 6420  
OY 6421 AAAGCTAGGGAAGGCTTGGGATGGGAGGCTTCTTCCAGGCCAGTAGATGAGAGCTGGTT 6480  
|||||  
Db 6421 AAAGCTAGGGAAGGCTTGGGATGGGAGGCTTCTTCCAGGCCAGTAGATGAGAGCTGGTT 6480  
OY 6481 AGCAGTGTGGCAGCTTCTCTCTGCTGTCAATAGCTATCCATTCACATCCATCCATCCAT 6540  
|||||  
Db 6481 AGCAGTGTGGCAGCTTCTCTCTGCTGTCAATAGCTATCCATTCACATCCATCCATCCATCCAT 6540  
OY 6541 ACACCCACCCATCCATTTATGACACCATCTTCCATCCATCCATCCATCCATCCATCCATCCAT 6600  
|||||  
Db 6541 ACACCCACCCATCCATTTATGACACCATCTTCCATCCATCCATCCATCCATCCATCCATCCAT 6600  
OY 6601 CCAGGCATCCATCCATCCATCCATCCATCCATCCATCCATCCATCCATCCATCCATCCAT 6660  
|||||  
Db 6601 CCAGGCATCCATCCATCCATCCATCCATCCATCCATCCATCCATCCATCCATCCATCCAT 6660  
OY 6661 CATTTATCCACAGAGAACTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 6720  
|||||  
Db 6661 CATTTATCCACAGAGAACTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 6720  
OY 6721 AGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 6780  
|||||  
Db 6721 AGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 6780  
OY 6781 TGGCAATCGGAGAAAGGTTTGGGCTGTGTTCTCTCCACAGCTGGGTTCTGGGGA 6840  
|||||  
Db 6781 TGGCAATCGGAGAAAGGTTTGGGCTGTGTTCTCTCCACAGCTGGGTTCTGGGGA 6840  
OY 6841 TTGAACTCAAAATATTCGGGCTGTGCAAGTGTCTTACACAGCCCAATTTTGGTGACA 6900  
|||||  
Db 6841 TTGAACTCAAAATATTCGGGCTGTGCAAGTGTCTTACACAGCCCAATTTTGGTGACA 6900  
OY 6901 CATCATTTATTATTAGAAAGCATCTTATGATGATGATGATGATGATGATGATGATGATGATG 6960  
|||||  
Db 6901 CATCATTTATTATTAGAAAGCATCTTATGATGATGATGATGATGATGATGATGATGATGATG 6960  
OY 6961 CCACGGATGACCTTTAACTCTGCTCTTCCAGGCTCCACCCGAGTCTAGGTTTACAGGT 7020  
|||||  
Db 6961 CCACGGATGACCTTTAACTCTGCTCTTCCAGGCTCCACCCGAGTCTAGGTTTACAGGT 7020  
OY 7021 GTTCAACCTGGTGAATGCTTTTATCCAGACACTCTGTGGGGGGGGGGGGGGGGGGGGGGGATC 7080  
|||||  
Db 7021 GTTCAACCTGGTGAATGCTTTTATCCAGACACTCTGTGGGGGGGGGGGGGGGGGGGGGGGATC 7080  
OY 7081 CCTGAGTTGGAGGCCAGTTGGTCTACAGAGTTTACAGATACCTGGGGCTATACAGAGGA 7140  
|||||  
Db 7081 CCTGAGTTGGAGGCCAGTTGGTCTACAGAGTTTACAGATACCTGGGGCTATACAGAGGA 7140

OY 7141 ACCCTATCCCAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAA 7200  
|||||  
Db 7141 ACCCTATCCCAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAA 7200  
OY 7201 TTAGAGATATTTAGTAGGATAGTAGGCTGTGGTAGGAGAGATCATGCTTTCTTTGTATT 7260  
|||||  
Db 7201 TTAGAGATATTTAGTAGGATAGTAGGCTGTGGTAGGAGAGATCATGCTTTCTTTGTATT 7260  
OY 7261 ATATATGTAAGTACACAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 7320  
|||||  
Db 7261 ATATATGTAAGTACACAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 7320  
OY 7321 TATCTATCTATCTACCTACCTACCTACCTACCTACCTACCTACCTACCTACCTACCTACCTAC 7380  
|||||  
Db 7321 TATCTATCTATCTACCTACCTACCTACCTACCTACCTACCTACCTACCTACCTACCTACCTAC 7380  
OY 7381 CTGCTTTGACTGTAATGCTCTGATTTCTGGGTCACCTTCAACCCCTAGTGTGGGTTT 7440  
|||||  
Db 7381 CTGCTTTGACTGTAATGCTCTGATTTCTGGGTCACCTTCAACCCCTAGTGTGGGTTT 7440  
OY 7441 ACCACACCCAGACATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 7500  
|||||  
Db 7441 ACCACACCCAGACATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 7500  
OY 7501 GGACTAGGGTCTTGTGATGCTATGACCAAGCTCTGCCCAGAGAGTGCAGTCCAGTCC 7560  
|||||  
Db 7501 GGACTAGGGTCTTGTGATGCTATGACCAAGCTCTGCCCAGAGAGTGCAGTCCAGTCCAGTCC 7560  
OY 7561 CCATTTTGTTCAGGTGATCTGTGACAGTTGTCAATTTGCGACGCTATGTACTCTCTC 7620  
|||||  
Db 7561 CCATTTTGTTCAGGTGATCTGTGACAGTTGTCAATTTGCGACGCTATGTACTCTCTC 7620  
OY 7621 CACCTCCAGTCCAGACATTTCTGGTCAATCCAGAGGGGGGGGCACTGTGCTCAGCA 7680  
|||||  
Db 7621 CACCTCCAGTCCAGACATTTCTGGTCAATCCAGAGGGGGGGGCACTGTGCTCAGCA 7680  
OY 7681 GTGCCCTGTCCCTGCTCTGACAGCTACATATTTGCTGTGCTGCTGCTGCTGCTGCTGCTGCT 7740  
|||||  
Db 7681 GTGCCCTGTCCCTGCTCTGACAGCTACATATTTGCTGTGCTGCTGCTGCTGCTGCTGCTGCT 7740  
OY 7741 GGAATGGCTTCCTGTGATTTCTTTATGAGGCGGCCCTTATCTTACACAGTTTGTGTTG 7800  
|||||  
Db 7741 GGAATGGCTTCCTGTGATTTCTTTATGAGGCGGCCCTTATCTTACACAGTTTGTGTTG 7800  
OY 7801 GGCCATGTGCACTGCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCT 7860  
|||||  
Db 7801 GGCCATGTGCACTGCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCT 7860  
OY 7861 TGTGTGATTAACACACTTCTGTTCATTTACTGATGAAATTTGTGGCCACCCCCAC 7920  
|||||  
Db 7861 TGTGTGATTAACACACTTCTGTTCATTTACTGATGAAATTTGTGGCCACCCCCAC 7920  
OY 7921 CCTTTTTTTTTTATTTGAGACAAGGCTTTTCTGTGTATCTTGCATCTTGGCTGTCT 7980  
|||||  
Db 7921 CCTTTTTTTTTTATTTGAGACAAGGCTTTTCTGTGTATCTTGCATCTTGGCTGTCT 7980  
OY 7981 GAGCTACGCTGTAGACCAAGGCTGTGAGGCTGTGCTCCACTTTGACACTCCTGTGAC 8040  
|||||  
Db 7981 GAGCTACGCTGTAGACCAAGGCTGTGAGGCTGTGCTCCACTTTGACACTCCTGTGAC 8040  
OY 8041 AGAGTAGCAGTGAACCTTCAAGACAAATTTCTGTTGGTTTACATTTGAGT 8100  
|||||  
Db 8041 AGAGTAGCAGTGAACCTTCAAGACAAATTTCTGTTGGTTTACATTTGAGT 8100  
OY 8101 TGTATCGGTATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 8160  
|||||  
Db 8101 TGTATCGGTATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 8160  
OY 8161 TGGGACAGAGAAACAAACCGATGTGCCATCTCAGATACATACCATCTTGTATATGTA 8220  
|||||  
Db 8161 TGGGACAGAGAAACAAACCGATGTGCCATCTCAGATACATACCATCTTGTATATGTA 8220  
OY 8221 TGTATTATGATGATTTATTTAGTGTGCCAAGTATGACAGTATTTTGTGAGATTTTAC 8280  
|||||

	8221	TGTAATATATGATATGTTTATTAGTGTGCCCAAGATATGACGATATTTTGTGGAGTTTTCAC	8280
OY	8281	CTTCCCTTTGTGGGCTCTCCGCAATTAACCTACGCTCTCGGGCTATGTAGCAATGCTTTCA	8340
Db	8281	CTTCCCTTTGTGGGCTCTCCGCAATTAACCTACGCTCTCGGGCTATGTAGCAATGCTTTCA	8340
OY	8341	CTGCATATGAGCAATCTCGCTGCCCCCTGGTGGCACCCTCTCCCTTATTTTCCAGATGGGACTA	8400
Db	8341	CTGCATATGAGCAATCTCGCTGCCCCCTGGTGGCACCCTCTCCCTTATTTTCCAGATGGGACTA	8400
OY	8401	CGCACTGCACTGGCCTTAAGCTCAACCAATCATACAGATGGCTAGAGCAGGAGCACTCAG	8460
Db	8401	CGCACTGCACTGGCCTTAAGCTCAACCAATCATACAGATGGCTAGAGCAGGAGCACTCAG	8460
OY	8461	GGATATATCTGGGCTCTGCTCCCTCCACAGAGCTATAGATTTACAGGCATACATCTGCTGGAG	8520
Db	8461	GGATATATCTGGGCTCTGCTCCCTCCACAGAGCTATAGATTTACAGGCATACATCTGCTGGAG	8520
OY	8521	ATTTTAACTGTAATCTGAGATAGAGAGAGGCACTGTACCAATGAGAGGCTTCTTTTGT	8580
Db	8521	ATTTTAACTGTAATCTGAGATAGAGAGAGGCACTGTACCAATGAGAGGCTTCTTTTGT	8580
OY	8581	GTTTGGTTTGGTTTCTCTGCAATACAGTCAAGGCTGAATATGTAGCTGGGCTAC	8640
Db	8581	GTTTGGTTTGGTTTCTCTGCAATACAGTCAAGGCTGAATATGTAGCTGGGCTAC	8640
OY	8641	ATTAACATCTTGTCACAAAAGCCATATAGAGTAGAGAGGTGAGGCTTAAGAAGAGCCTT	8700
Db	8641	ATTAACATCTTGTCACAAAAGCCATATAGAGTAGAGAGGTGAGGCTTAAGAAGAGCCTT	8700
OY	8701	AAGCCGGCTGTGATAGACACACAGGATAGCCTGACATATATAGCAAGACCTTGTTCAAA	8760
Db	8701	AAGCCGGCTGTGATAGACACACAGGATAGCCTGACATATATAGCAAGACCTTGTTCAAA	8760
OY	8761	ACATGGAGGGAGGGGCTATGTTTAAAGTCGCGGCTGTATACAGGCACTAAGGAGGCCAA	8820
Db	8761	ACATGGAGGGAGGGGCTATGTTTAAAGTCGCGGCTGTATACAGGCACTAAGGAGGCCAA	8820
OY	8821	TGTAGACATTTGACTAAGAAAGGATCATCATCAAGCCGGGTGGGCAAGGATAGAGTTGG	8880
Db	8821	TGTAGACATTTGACTAAGAAAGGATCATCATCAAGCCGGGTGGGCAAGGATAGAGTTGG	8880
OY	8881	ACTACAGTGTCAAGACCCCATATAGAAAGCCAGTTTCCCTTCTCTGGGCTCAAGC	8940
Db	8881	ACTACAGTGTCAAGACCCCATATAGAAAGCCAGTTTCCCTTCTCTGGGCTCAAGC	8940
OY	8941	CTGGCTGAGAGGCACTGCTCTCAACATGACCTTCTCTCTAGGCTCGTCCACCATG	8995
Db	8941	CTGGCTGAGAGGCACTGCTCTCAACATGACCTTCTCTCTAGGCTCGTCCACCATG	8995
RESULT 2			
AAAA9605			
AAAA9605 standard; DNA; 8993 BP.			
XX	AAA49605:		
AC	AAA49605:		
XX	19-DEC-2000 (first entry)		
DT	19-DEC-2000 (first entry)		
XX	Mouse villin gene regulatory elements and partial sequence.		
DE	Mouse: villin; intestinal epithelial cell;		
XX	uro-genital tract epithelial cell; tumour; ds.		
KW	Mouse: villin; intestinal epithelial cell;		
XX	uro-genital tract epithelial cell; tumour; ds.		
XX	Mus sp.		
OS	Mus sp.		
XX			
XX			
FT	Key	Location/Qualifiers	
FT	misc_signal	3442	
FT		/*tag= a	
FT		/note= "transcriptional start site"	
FT	exon	3442..3487	
FT		/*tag= b	

FT	/number=1
FT	3488..8981
FT	//tag=c
FT	/number=1
FT	8982..8995
FT	//tag=d
FT	/number=2
FT	8992
FT	//tag=e
FT	/note="translational start site"
XX	
PN	WO200034493-A2.
PD	15-JUN-2000.
XX	
PF	09-DEC-1999; 99WO-EP09782.
XX	
PR	09-DEC-1998; 98WO-EP08009.
XX	
PA	(CNRS-) CENT NAT RECH SCI.
XX	(CURIE-) INST CURIE.
PI	Pinto D, Robine S, Jaisser F, Louvard D, Niewoehner J;
XX	
DR	WPI: 2000-423434/36.
PT	Novel nucleotide sequence derived from mouse villin gene for targeted
PT	expression of transgenes in immature and differentiated epithelial
PT	cells of intestine or urogenital tracts -
XX	
PS	Claim 3; Fig 6; 52pp; English.
XX	
CC	The present sequence consists of the regulatory region and the first
CC	exon and intron of the murine villin gene. This gene is expressed in the
CC	epithelial cells of the intestine and uro-genital tracts. Its promoter
CC	sequence can be used in the targeted expression of exogenous genes in
CC	these places, which may, for example, be useful in the treatment of
CC	tumours.
XX	
SQ	Sequence 8993 BP; 2274 A; 2105 C; 2257 G; 2355 T; 2 other;
	Query Match 96.2%; Score 8651; DB 21; Length 8993;
	Best Local Similarity 100.0%; Pred. No. 0;
	Matches 8991; Conservative 0; Mismatches 2; Indels 2; Gaps 2
QY	1 GATCTGTCACCAAGACACTGTGGTCCAGCACTGGGAGGTGGAGGAGGTCA 60
DB	1 GATCTGGGCACCAGAAGACACTGTGGTCCAGCACTGGGAGGTGGAGGAGGTCA 60
QY	61 GAACTTTAAGGTCATTCCCTTGTTACATAGCAAGGTTTAGCAGCTTAGCATGAA 120
DB	61 GAAGTTTAAAGTCAATCCTTGTTACATAGCAAGGTTTAGCAGCTTAGCATGAAA 120
QY	121 CCTTGTGTTGTGTGTGTGTGTTTAAAGCATTAATAAATAATCACATAAGAGGTTGG 180
DB	121 CCTTGTGTTGTGTGTGTGTGTTTAAAGCATTAATAAATAATCACATAAGAGGTTGG 180
QY	181 CAGTGGTGGACAACCTTTAATTCAGATATTCCAGGAGGCGAAGCAGAGATCTCTGT 240
DB	181 CAGTGGTGGCGACGAACCTTTAATTCAGATATTCCAGGAGGCGAAGCAGAGATCTCTGT 240
QY	241 GAGTTTCGAAGCAGCCTAGTGTGCAAAAGCTAGTTTCAGAGATGGCAAGGGCTAACAGAGA 300
DB	241 GAGTTTCGAAGCAGCCTAGTGTGCAAAAGCTAGTTTCAGAGATGGCAAGGGCTAACAGAGA 300
QY	301 AACCTGTCTCATAAACCAAAGTAGTAGTAGTAGTAATGCCATAGAGAAATTGGA 360
DB	301 AACCTGTCTCATAAACCAAAGTAGTAGTAGTAGTAATGCCATAGAGAAATTGGA 360
QY	361 GTCCATTTCAGATGAGCAATCCTTAAGATATCTCTGACCCAGTAAGATATGTGCA 420
DB	361 GTCCATTTCAGATGAGCAATCCTTAAGATATCTCTGACCCAGTAAGATATGTGCA 420

QY 421 TGGGAAAGGGATGGGACTGTCCTAGATTAAAAAGTGCTGAGCGATGCTATTCTCAA 480  
Db 421 TGGGAAAGGGATGGGACTGTCCTAGATTAAAAAGTGCTGAGCGATGCTATTCTCAA 480  
QY 481 TTTTGATTCCATATGAAAAAGGCTGTATAGGCCCAAGAGAAAGTGAAGTGGGACTCTGACT 540  
Db 481 TTTTGATTCCATATGAAAAAGGCTGTATAGGCCCAAGAGAAAGTGAAGTGGGACTCTGACT 540  
QY 541 GAAGAGGTAGCGGCTTTATTAACACTGGCACTTATTAACACTTATTAACACTGGCAGAG 600  
Db 541 GAAGAGGTAGCGGCTTTATTAACACTGGCACTTATTAACACTTATTAACACTGGCAGAG 600  
QY 601 CGTTCAAGTTTGAAGATCACTTTCAACACACAGAACAGAAAGTGTCTGCTGCTCAGC 660  
Db 601 CGTTCAAGTTTGAAGATCACTTTCAACACACAGAACAGAAAGTGTCTGCTGCTCAGC 660  
QY 661 GTAAGCAGCACTGGCTGCGAGAAAGTGTATTTAGTGAAGAGTACCTTCAACATATCTTT 720  
Db 661 GTAAGCAGCACTGGCTGCGAGAAAGTGTATTTAGTGAAGAGTACCTTCAACATATCTTT 720  
QY 721 GCACCTATACATACACAGTGTCAAAATGTGCTAACTCCCTAGTCCACAGATGGCTGTACA 780  
Db 721 GCACCTATACATACACAGTGTCAAAATGTGCTAACTCCCTAGTCCACAGATGGCTGTACA 780  
QY 781 CTCGTTTCTGCTTCCCATCTGGTTGACATTTGTGCAAAACAGAAATTAAGAAATGTGGT 840  
Db 781 CTCGTTTCTGCTTCCCATCTGGTTGACATTTGTGCAAAACAGAAATTAAGAAATGTGGT 840  
QY 841 ATTTATTTGTGCTGCTGAGACACATCCAGAGGCTTTTACATTTACAGGCATATGTTTAC 900  
Db 841 ATTTATTTGTGCTGCTGAGACACATCCAGAGGCTTTTACATTTACAGGCATATGTTTAC 900  
QY 901 TTAAGTGGCTACTTCTCCAAAGGTTTGAACACATTTGTTTATATTACTATTTTGTGT 960  
Db 901 TTAAGTGGCTACTTCTCCAAAGGTTTGAACACATTTGTTTATATTACTATTTTGTGT 960  
QY 961 GCATGAGGTAGGCATGTATACGTATGTATAGAGTCAATGCATGGCTGCTACCTCAA 1020  
Db 961 GCATGAGGTAGGCATGTATACGTATGTATAGAGTCAATGCATGGCTGCTACCTCAA 1020  
QY 1021 ATCATTTGCAATCCCGAGAGTGAAGTACCGAGGTTGTAGTTTATATGAGGACATC 1080  
Db 1021 ATCATTTGCAATCCCGAGAGTGAAGTACCGAGGTTGTAGTTTATATGAGGACATC 1080  
QY 1081 GGAAGCCAAAGGCTGGTCTCTGCAAGAGCAGCAGTGGCTTAACATGGGACACACTCT 1140  
Db 1081 GGAAGCCAAAGGCTGGTCTCTGCAAGAGCAGCAGTGGCTTAACATGGGACACACTCT 1140  
QY 1141 CTAGGCTTAAGGTAACTTTTATTTTAAAAATATATATCTCAGCCGGGTGTGGTGC 1200  
Db 1141 CTAGGCTTAAGGTAACTTTTATTTTAAAAATATATATCTCAGCCGGGTGTGGTGC 1200  
QY 1201 ACACGCTTTAATCCAGACCTTGAGAGCTGAGGTAGGAATTTATACACAGGCCAG 1260  
Db 1201 ACACGCTTTAATCCAGACCTTGAGAGCTGAGGTAGGAATTTATACACAGGCCAG 1260  
QY 1261 CTGGGGTGCAGAGCTTGGCCCTGTTTTTTTGTGTTTTCTTTATGTGCACTGGTGTCTTA 1320  
Db 1261 CTGGGGTGCAGAGCTTGGCCCTGTTTTTTTGTGTTTTCTTTATGTGCACTGGTGTCTTA 1320  
QY 1321 CCTGCGTGTATGCTCGTGCAAGGGTGTAGATCCCTTGGAGCTGGAATTAAAGACAGTTG 1380  
Db 1321 CCTGCGTGTATGCTCGTGCAAGGGTGTAGATCCCTTGGAGCTGGAATTAAAGACAGTTG 1380  
QY 1381 TGTATCAGCTGCTGCTTACAGATGCTGGAATTTGAACCCAGGTGCTCCCTGAGAGAGAGC 1440  
Db 1381 TGTATCAGCTGCTGCTTACAGATGCTGGAATTTGAACCCAGGTGCTCCCTGAGAGAGAGC 1440  
QY 1441 AGTGTCTTAATCTTGAAGCACCCTTCAACCTCTTTTGAAGACTCTTAACCTTTTG 1500  
Db 1441 AGTGTCTTAATCTTGAAGCACCCTTCAACCTCTTTTGAAGACTCTTAACCTTTTG 1500  
QY 1501 TGTAAATGTGGAACTGAGTGTGATCTTGACATTACCAAGTGTGTGCTGCGCTGTAGCATCA 1560

Db 1501 TGTAAATGTGGAACTGAGTGTGATCTTGACATTACCAAGTGTGTGCTGCGCTGTAGCATCA 1560  
QY 1561 CTGAGCCCGTACCACACAGCATGTAGTATACATTTTAAGGCAAAACACTTAACAAATGACA 1620  
Db 1561 CTGAGCCCGTACCACACAGCATGTAGTATACATTTTAAGGCAAAACACTTAACAAATGACA 1620  
QY 1621 ATAGTTGATAGAGTTTGAATATATAGTCCAGGCTAATTTGTTTACGCTGACCTTGTCTGC 1680  
Db 1621 ATAGTTGATAGAGTTTGAATATATAGTCCAGGCTAATTTGTTTACGCTGACCTTGTCTGC 1680  
QY 1681 TTAGCATGTGCTGTGAGAAAGATGAAAAATGAAGACTTGAAGTGTCTGGAACCCACA 1740  
Db 1681 TTAGCATGTGCTGTGAGAAAGATGAAAAATGAAGACTTGAAGTGTCTGGAACCCACA 1740  
QY 1741 GAGGCAAGCGAGAACCCACTCCCTGGAAGTGTCTGTGAGCTTCACTATCACTTACAT 1800  
Db 1741 GAGGCAAGCGAGAACCCACTCCCTGGAAGTGTCTGTGAGCTTCACTATCACTTACAT 1800  
QY 1801 AATAGTTACATGAT 1860  
Db 1801 AATAGTTACATGAT 1860  
QY 1861 GAGATGGCTCAGCTTCCAGAGGACCTTGTCTTGTGCAAGAGACCTAGATTGAGTTCC 1920  
Db 1861 GAGATGGCTCAGCTTCCAGAGGACCTTGTGTCTGCAAGAGGACCTAGATTGAGTTCC 1920  
QY 1921 AGGACTCATATGTTGGCTCAGACACATCTGTAAATCCAGTCCAGAGAGGTTCCACACCT 1980  
Db 1921 AGGACTCATATGTTGGCTCAGACACATCTGTAAATCCAGTCCAGAGAGGTTCCACACCT 1980  
QY 1981 CTTCGCGCTCCACAGGACACACATATATATATATATATATATATATATATATATATAT 2040  
Db 1981 CTTCGCGCTCCACAGGACACACATATATATATATATATATATATATATATATATATAT 2040  
QY 2041 CATACACATTAATTAATTAAGAAACTTAAGAAAGTGTGATGTGTGTTGTTGTTGCT 2100  
Db 2041 CATACACATTAATTAATTAAGAAACTTAAGAAAGTGTGATGTGTGTTGTTGTTGCT 2100  
QY 2101 TACACATGCTGATTTGAAGACATGTACAGGACACACTGGAAGAGGTTCTGGGGCTGGAG 2160  
Db 2101 TACACATGCTGATTTGAAGACATGTACAGGACACACTGGAAGAGGTTCTGGGGCTGGAG 2160  
QY 2161 AGATGGCTCAGCGGTTAAGAGCATGACTGCTTCCGAAGGAGTCTGATTTCAAT 2220  
Db 2161 AGATGGCTCAGCGGTTAAGAGCATGACTGCTTCCGAAGGAGTCTGATTTCAAT 2220  
QY 2221 CCTAGCAACCATGCTGCTCACAACCATCATTAATGAGATCTGACACCTCTTGTGT 2280  
Db 2221 CCTAGCAACCATGCTGCTCACAACCATCATTAATGAGATCTGACACCTCTTGTGT 2280  
QY 2281 GCATCTGAAGACGCTGCAAGAGCTACAGTGTACTTGTATTAATTAATTAATCTTTT 2340  
Db 2281 GCATCTGAAGACGCTGCAAGAGCTACAGTGTACTTGTATTAATTAATTAATCTTTT 2340  
QY 2341 TTTTAAAAAATGAAGGAGTGTGAGACACCTCAAAAGAGATTTAGAGTACTCAGC 2400  
Db 2341 TTTTAAAAAATGAAGGAGTGTGAGACACCTCAAAAGAGATTTAGAGTACTCAGC 2400  
QY 2401 GGTGATTTATCTATCTGAGATTTTCTCTTCTGCTTGGCTTGGCACTGGGTGACAGAGC 2460  
Db 2401 GGTGATTTATCTATCTGAGATTTTCTCTTCTGCTTGGCTTGGCACTGGGTGACAGAGC 2460  
QY 2461 CCCCTTTTCAATTCACAAAGAGGCTCTCATATTTTCTGACAAACAAAGCACCCTGCACT 2520  
Db 2461 CCCCTTTTCAATTCACAAAGAGGCTCTCATATTTTCTGACAAACAAAGCACCCTGCACT 2520  
QY 2521 ATGTTTACTGTCTTGTCTACTATGTAGCAGCGCAGCGCGCGCGCAGACACACACAC 2580  
Db 2521 ATGTTTACTGTCTTGTCTACTATGTAGCAGCGCAGCGCGCGCGCAGACACACACAC 2580  
QY 2581 ACACACACACACACACACACACACACATTCAGTCTCCAGAGCTCTTGGAGAGTCA 2640

D 2581 ACACACACACACACACACACACACATTCAGTCTCCAGAGCTTTGGAGGTCA 2640  
OY 2641 AGAAGAGGCTGCCCTCAAAACAGATCTTCATCTTTCCCTCTAAAGAGACAGATTC 2700  
D 2641 AGAAGAGGCTGCCCTCAAAACAGATCTTCATCTTTCCCTCTAAAGAGACAGATTC 2700  
OY 2701 AAGGTGGCAGAGATCTACAGGGGGCAGAGCCAGGAGGGGGAACAGAGCCATGGTTCC 2760  
D 2701 AAGGTGGCAGAGATCTACAGGGGGCAGAGCCAGGAGGGGGAACAGAGCCATGGTTCC 2760  
OY 2761 AGAGACCTACAGAGAGGGGAGCAAGAGATCCCCAGGTCCAGGGGAGGGAGGGAGG 2820  
D 2761 AGAGACCTACAGAGAGGGGAGCAAGAGATCCCCAGGTCCAGGGGAGGGAGGGAGG 2820  
OY 2821 CCCTTGTTCCAGAGAGAGAGGCGAGAGCCAGAGAGGTTCAAAGGACAGAGTTATGCA 2880  
D 2821 CCCTTGTTCCAGAGAGAGAGGCGAGAGCCAGAGAGGTTCAAAGGACAGAGTTATGCA 2880  
OY 2881 GCTCATAAAGTGAAGTCTGTGGCTCACTCAGAAAGAGAGAAAGAGAGAGCCCTTGT 2940  
D 2881 GCTCATAAAGTGAAGTCTGTGGCTCACTCAGAAAGAGAGAAAGAGAGAGCCCTTGT 2940  
OY 2941 GCCCAGTGAAGGAGGTATGCTGAGTAGAGAGATCTGCAAGGGGTCCAGAGAGCCAC 3000  
D 2941 GCCCAGTGAAGGAGGTATGCTGAGTAGAGAGATCTGCAAGGGGTCCAGAGAGCCAC 3000  
OY 3001 CTGTCTGTCCAG 3060  
D 3001 CTGTCTGTCCAG 3060  
OY 3061 CAAGACCCCAAGAGATCTACATCCCATATCAGAGAGAGAGAGAGAGAGAGAGAG 3120  
D 3061 CAAGACCCCAAGAGATCTACATCCCATATCAGAGAGAGAGAGAGAGAGAGAGAG 3120  
OY 3121 CCCCCAGTCCCGTCCAGTCTCTAGAGGCTGAGAGAGAGAGAGAGAGAGAGAGAG 3180  
D 3121 CCCCCAGTCCCGTCCAGTCTCTAGAGGCTGAGAGAGAGAGAGAGAGAGAGAGAG 3180  
OY 3181 CTACCTGAGGTAG 3240  
D 3181 CTACCTGAGGTAG 3240  
OY 3241 GCCAAGGGGGGGGAG 3300  
D 3241 GCCAAGGGGGGGGAG 3300  
OY 3301 AGGAGCTGGGACCTACTTCCCAACCCCATCATCTCTCTGGGGCCCTATCTCC 3360  
D 3301 AGGAGCTGGGACCTACTTCCCAACCCCATCATCTCTCTGGGGCCCTATCTCC 3360  
OY 3361 CTATATGCTGAG 3420  
D 3361 CTATATGCTGAG 3420  
OY 3421 GTTCCTGACAG 3480  
D 3421 GTTCCTGACAG 3480  
OY 3481 CAGACAGGTAAAGGCAATTGGGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3540  
D 3481 CAGACAGGTAAAGGCAATTGGGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3540  
OY 3541 GGTCCTGCTCTCTCTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3600  
D 3541 GGTCCTGCTCTCTCTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3600  
OY 3601 TGAGGTAAAGGTGCTGAACTCTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3660  
D 3601 TGAGGTAAAGGTGCTGAACTCTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3660  
OY 3661 TCAATGAAG 3720  
D 3661 TCAATGAAG 3720

OY 3721 ATTCCTGGGAGCAAGGTGGCTCTGGAGCTCAGATTTCTACAAATTAATCAGACAGT 3780  
D 3721 ATTCCTGGGAGCAAGGTGGCTCTGGAGCTCAGATTTCTACAAATTAATCAGACAGT 3780  
OY 3781 CCTGAGACTTGAGCTCCGCTGCTATTTACTACTTCTCTGGCTGCTCATTTCTGTGT 3840  
D 3781 CCTGAGACTTGAGCTCCGCTGCTGCTATTTACTACTTCTCTGGCTGCTCATTTCTGTGT 3840  
OY 3841 TCATGTCTTACACATCTGAATGTTTCTTGTGTGACACATTTCCCTGACACTCTGGGA 3900  
D 3841 TCATGTCTTACACATCTGAATGTTTCTTGTGTGACACATTTCCCTGACACTCTGGGA 3900  
OY 3901 GGTCGTATCTTGGCACATGTATCTGAGATGTAACTGACAGCCACAGAGAGAGGGG 3960  
D 3901 GGTCGTATCTTGGCACATGTATCTGAGATGTAACTGAGAGCCACAGAGAGAGGGG 3960  
OY 3961 AGAGTCAAGAGACTGTGCTAGGGCCCTTTAGGCTTGACATCAACCCCTTTCTAGAAAT 4020  
D 3961 AGAGTCAAGAGACTGTGCTAGGGCCCTTTAGGCTTGACATCAACCCCTTTCTAGAAAT 4020  
OY 4021 GGCCCTTCATTTTTCGTTACCATGATCTATTTATATCAGAGTGGGACAGTAAAGCA 4080  
D 4021 GGCCCTTCATTTTTCGTTACCATGATCTATTTATATCAGAGTGGGACAGTAAAGCA 4080  
OY 4081 AACCTGCCAGAGAGTTTGGAGCTCATCAGACCAAGGTTATCTGCTAGAAATCCCTCTG 4140  
D 4081 AACCTGCCAGAGAGTTTGGAGCTCATCAGACCAAGGTTATCTGCTAGAAATCCCTCTG 4140  
OY 4141 TCACCTGAGGTGGGAGAGATGCTCTGAGGGGCTCCAGAGTCTTGTTAGCAGAGAGGT 4200  
D 4141 TCACCTGAGGTGGGAGAGATGCTCTGAGGGGCTCCAGAGTCTTGTTAGCAGAGAGGT 4200  
OY 4201 ATCTCTTGTATAGGAGCATGACCTAGTATGTTTACTATGATCTCTCCAGTTAAAG 4260  
D 4201 ATCTCTTGTATAGGAGCATGACCTAGTATGTTTACTATGATCTCTCCAGTTAAAG 4260  
OY 4261 CTGGNACTAAACCAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4320  
D 4261 CTGGNACTAAACCAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4320  
OY 4321 CTGGNACTAAACCAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4380  
D 4321 CTGGNACTAAACCAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4380  
OY 4381 GCCCAGAGAGTCCGTTCCCTAGTTCACAATGCCAGTATGAGTGTAGCTACTATGGCTG 4440  
D 4381 GCCCAGAGAGTCCGTTCCCTAGTTCACAATGCCAGTATGAGTGTAGCTACTATGGCTG 4440  
OY 4441 TGAGTTGGTACTACAGAGATGAGTGTGATGTGTGTGTGTGTGTGTGTGTGTGTGT 4500  
D 4441 TGAGTTGGTACTACAGAGATGAGTGTGATGTGTGTGTGTGTGTGTGTGTGTGTGT 4500  
OY 4501 TTGGAGAGCTGAAGAGAGAGATGCTATGTTTGAAGCCAGCTGATGAGAGCA 4560  
D 4501 TTGGAGAGCTGAAGAGAGAGATGCTATGTTTGAAGCCAGCTGATGAGAGCA 4560  
OY 4561 GACTTGTCTTTAAGAAAAAATGAAGCCAGAGAGTGTGGCAGACGCTTTAATCCCA 4620  
D 4561 GACTTGTCTTTAAGAAAAAATGAAGCCAGAGAGTGTGGCAGACGCTTTAATCCCA 4620  
OY 4621 GCATCTGGAGAGCAGAGAGAGAGATTTCTGAGTTCAGAGCCAGAGCTGTATAGAGT 4680  
D 4621 GCATCTGGAGAGCAGAGAGAGAGATTTCTGAGTTCAGAGCCAGAGCTGTATAGAGT 4680  
OY 4681 GAGTTCAG 4740  
D 4681 GAGTTCAG 4740  
OY 4741 AACAAACAAACAAACAAACAAACAAACAAACAAACAAACCTCTCATCTCTCTC 4800  
D 4741 AACAAACAAACAAACAAACAAACAAACAAACAAACCTCTCATCTCTCTC 4800

QY	4801	TAGCGTGTCTGTCTAGTGGTAGAGTTTGGGCACTTCAACATTATATATAATAGGCC	4860
Db	4801	TAGCGTGTCTGTCTAGTGGTAGAGTTTGGGCACTTCAACATTATATATAATAGGCC	4860
QY	4861	TTTTATACACGCTAGAGACGAGAAAGCTTACGTCTGGACACAGTGGGACCTGAGA	4920
Db	4861	TTTTATACGTGTAGAGACGAGAAAGTTTCACTCTGGGACACAGTGGGACCTGAGA	4920
QY	4921	AAGTACTCTTGGCCAGCCCAAAAATTCCTGGGAAGGCTTCTCGGAGGAAGTGTGCCGAT	4980
Db	4921	AAGTACTCTTGGCCAGCCCAAAAATTCCTGGGAAGGCTTCTCGGAGGAAGTGTGCCGAT	4980
QY	4981	CAGACTACTGTTCTAGAAAGCAGAAAGAGGCTTGGAAATGTGTGTGCACAGACTT	5040
Db	4981	CAGACTACTGTTCTAGAAAGCAGAAAGAGGCTTGGAAATGTGTGTGCACAGACTT	5040
QY	5041	GGAAACAGAAAGCAGAGAGGGGAGGCATCCAAAGATTCGAAACATGTAGCTACCTTTGGT	5100
Db	5041	GGAAACAGAAAGCAGAGAGGGGAGGCATCCAAAGATTCGAAACATGTAGCTACCTTTGGT	5100
QY	5101	TCTCTGGGTGCAAAAGTGTCCCCCAGAGGATAGGGGTGTAGAAAGGGGACAGAGGCTGAGCC	5160
Db	5101	TCTCTGGGTGCAAAAGTGTCCCCCAGAGGATAGGGGTGTAGAAAGGGGACAGAGGCTGAGCC	5160
QY	5161	AATGACTTCAAGTTGAGGGAGACATCCAGGCCAGGGTCTTGCTGGCAAGCTAAAGAAATG	5220
Db	5161	AATGACTTCAAGTTGAGGGAGACATCCAGGCCAGGGTCTTGCTGGCAAGCTAAAGAAATG	5220
QY	5221	AGAGCCCTTAACCCCTCCCTGAAGTTTGGGGACACAGAGAGCTGAGAGATCTCTCTA	5280
Db	5221	AGAGCCCTTAACCCCTCCCTGAAGTTTGGGGAGACAGAGAGCTGAGAGATCTCTCTA	5280
QY	5281	GGGTGAAGAGAGAGTATGCTCTGACCAATGGCTGTGGAGCAGAAACAGTGTGACACAG	5340
Db	5281	GGGTGAAGAGAGAGTATGCTCTGACCAATGGCTGTGGAGCAGAAACAGTGTGACACAG	5340
QY	5341	TTACCCCTCAGAACCCAGCCATCCCTCTTGGCTCTAAGAGAGCTGGGCCCTTCTGTCT	5400
Db	5341	TTACCCCTCAGAACCCAGCCATCCCTCTTGGCTCTAAGAGAGCTGGGCCCTTCTGTCT	5400
QY	5401	AAGATCTTACTTTTCTTCAGAGAGAGGCACAGCCTTGTCCCCCTCTGTGTGTCA	5460
Db	5401	AAGATCTTACTTTTCTTCAGAGAGAGGCACAGCCTTGTCCCCCTCTGTGTGTCA	5460
QY	5461	TAAACACCCCTGTGTGAACATTTAGTTATTTTACTGTCACTTTTGTGTCACAGACAGTCCA	5520
Db	5461	TAAACACCCCTGTGTGAACATTTAGTTATTTTACTGTCACTTTTGTGTCACAGACAGTCCA	5520
QY	5521	TCTGTAGACCTGTCTCCTTAACCTCACCAGAGTATGGCCCATTTCTCACCCAGAGAG	5580
Db	5521	TCTGTGTAGACCTGTCTCCTTAACCTCACCAGAGTATGGCCCATTTCTCACCCAGAGAG	5580
QY	5581	TGCAGAAAGAGCCTTAGAGAAAGGGTAAACGTAACAAAGATGGCCAGATATAAACAAA	5640
Db	5581	TGCAGAAAGAGCCTTAGAGAAAGGGTAAACGTAACAAAGATGGCCAGATATAAACAAA	5640
QY	5641	ACTACTATCTTTGTGACCCAAATTTGTTTCTGTGACCAAGAGAGGGGGTGTGTAGTGTAT	5700
Db	5641	ACTACTATCTTTGTGACCCAAATTTGTTTCTGTGACCAAGAGAGGGGGTGTGTAGTGTAT	5700
QY	5701	GT	5760
Db	5701	GT	5760
QY	5761	CTTGGGGGACCTTTCATCTTAAGAAATATCGATATATGGCCCATGTCAACAGGGGTAT	5820
Db	5761	CTTGGGGGACCTTTCATCTTAAGAAATATCGATATATGGCCCATGTCAACAGGGGTAT	5820
QY	5821	TGGGAGAGGTAGGCTTCTGCAAAACAGTAAAGGCCCAAGATGAGATTTGGTGGCTGAA	5880
Db	5821	TGGGAGAGGTAGGCTTCTGCAAAACAGTAAAGGCCCAAGATGAGATTTGGTGGCTGAA	5880
QY	5881	TCACCAAGGGGACGGCTGATCAGATGTGACAGAACATCACAAAGTAAAGCCACCCCTGTGGG	5940

[illegible]

Db 6960 CCACGATGACCTTTAACTCCTGCTTCCAGCCTCCACCCGAGCTGTAGGTTTACAGGT 7019  
 Qy 7021 GTTCAACTGGTGAATGCTTTAAATCCAGCAGCTGTGGGGGGGGGGGAGGAGCGGATC 7080  
 Db 7020 GTTCAACTGGTGAATGCTTTAAATCCAGCAGCTGTGGGGGGGGGGGAGGAGCGGATC 7079  
 Qy 7081 CTTGAGTGGAGGCGGAGTTGGTCTACAGAGTTTACAGATACCTGGGGGCTATACAGGAA 7140  
 Db 7080 CTTGAGTGGAGGCGGAGTTGGTCTACAGAGTTTACAGATACCTGGGGGCTATACAGGAA 7139  
 Qy 7141 ACCCTATCCCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 7200  
 Db 7140 ACCCTATCCCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 7199  
 Qy 7201 TTAGAGGATATAGTAGGAGTAGAGGCGTGTGAGGAGAGTACGCTTTCTTTTGTATT 7260  
 Db 7200 TTAGAGGATATAGTAGGAGTAGAGGCGTGTGAGGAGAGTACGCTTTCTTTTGTATT 7259  
 Qy 7261 ATATAGTAAAGTACTACAAAGATGCAATATCTATCTATCTATCTATCTATCTATCTATC 7320  
 Db 7260 ATATAGTAAAGTACTACAAAGATGCAATATCTATCTATCTATCTATCTATCTATCTATC 7319  
 Qy 7321 TATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATC 7380  
 Db 7320 TATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATC 7379  
 Qy 7381 CTGCTTTGACTGTGAATGCTCTAATTTCTGGGTCACCTCTACCCCTAGTGTGGGTTT 7440  
 Db 7380 CTGCTTTGACTGTGAATGCTCTAATTTCTGGGTCACCTCTACCCCTAGTGTGGGTTT 7439  
 Qy 7441 ACCAACCACGACATTTATTTATTTTGTATTATTTATTTATTTATTTATTTATTTATTTAT 7500  
 Db 7440 ACCAACCACGACATTTATTTATTTTGTATTATTTATTTATTTATTTATTTATTTATTTAT 7499  
 Qy 7501 GGAATCAGGGTCTTGTGATGCTAAGCAAGCTCTGTCACAGAGCTGACGCTCAATCC 7560  
 Db 7500 GGAATCAGGGTCTTGTGATGCTAAGCAAGCTCTGTCACAGAGCTGACGCTCAATCC 7559  
 Qy 7561 CCATTTTGTTCAGTACTGTGACAGTTGTGACATTTGTGACAGGCTATGTAGCTCTC 7620  
 Db 7560 CCATTTTGTTCAGTACTGTGACAGTTGTGACATTTGTGACAGGCTATGTAGCTCTC 7619  
 Qy 7621 CACCTCCAGTTCCAGAGCTTCTGTCATCCAGTGGGGGGGCACTGTGTGTCACCA 7680  
 Db 7620 CACCTCCAGTTCCAGAGCTTCTGTCATCCAGTGGGGGGGCACTGTGTGTCACCA 7679  
 Qy 7681 GTGCGCTGTCCCTGTCTGTCAGACCTAATTTGCTGTGGAAGCTTCAATGAATG 7740  
 Db 7680 GTGCGCTGTCCCTGTCTGTCAGACCTAATTTGCTGTGGAAGCTTCAATGAATG 7739  
 Qy 7741 GGATGCGTCTCTGTATCTTTTATGCTGGGCCCTTTATCTTGAACAGTTTGTGTG 7800  
 Db 7740 GGATGCGTCTCTGTATCTTTTATGCTGGGCCCTTTATCTTGAACAGTTTGTGTG 7799  
 Qy 7801 GGCATGTGTCACGTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCT 7860  
 Db 7800 GGCATGTGTCACGTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCT 7859  
 Qy 7861 TGTGTGATTAACCACTTTCTGTTTCAATTTAGATGGAATTTGTGTGCCCCACCCAC 7920  
 Db 7860 TGTGTGATTAACCACTTTCTGTTTCAATTTAGATGGAATTTGTGTGCCCCACCCAC 7919  
 Qy 7921 CCTTTTATTTTATTTGAGACAAGTCTTGTGTGTAATCTTGGCTGTCT 7980  
 Db 7920 CCTTTTATTTTATTTGAGACAAGTCTTGTGTGTAATCTTGGCTGTCT 7979  
 Qy 7981 GAGCTCAGCTGTAGACAGAGCTGTGAGCTGTCTTCCACTTTTGACACTCCGTGAAC 8040  
 Db 7980 GAGCTCAGCTGTAGACAGAGCTGTGAGCTGTCTTCCACTTTTGACACTCCGTGAAC 8039  
 Qy 8041 AGAGTACCATGAACCTTCAAGACAATTTTCTGTTTGGTTTGGTTTTCATTGTGTG 8100  
 Db 8040 AGAGTACCATGAACCTTCAAGACAATTTTCTGTTTGGTTTGGTTTTCATTGTGTG 8098

Qy 8101 TGTATGCGTGTATATGTGATGTTTGTGTCTTCAAGTGTCTACATGTGTGTACCTGTGTG 8160  
 Db 8099 TGTATGCGTGTATATGTGATGTTTGTGTCTTCAAGTGTCTACATGTGTGTACCTGTGTG 8158  
 Qy 8161 TGGACAGAGAACAAACCGATGTGCCATTTCTCTCAGATACATACGATCTTGTATATGTA 8220  
 Db 8159 TGGACAGAGAACAAACCGATGTGCCATTTCTCTCAGATACATACGATCTTGTATATGTA 8218  
 Qy 8221 TGTATATGTATGTATTTATTTAGTGTGCCAAGTATGAGGATATTTTGTGAGTTTTCAC 8280  
 Db 8219 TGTATATGTATGTATTTATTTAGTGTGCCAAGTATGAGGATATTTTGTGAGTTTTCAC 8278  
 Qy 8281 CTTCCTTGTGGGCTCTCCGATTAACCTAGCTCTCGGGCTAGTACGAATGCTTCA 8340  
 Db 8279 CTTCCTTGTGGGCTCTCCGATTAACCTAGCTCTCGGGCTAGTACGAATGCTTCA 8338  
 Qy 8341 CTGATATGACATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 8400  
 Db 8339 CTGATATGACATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 8398  
 Qy 8401 CGCAGTCACTGGCTTAAGCTCAACAGTATCCAGATGAGTGTGAGGAGGAGACTCAG 8460  
 Db 8399 CGCAGTCACTGGCTTAAGCTCAACAGTATCCAGATGAGTGTGAGGAGGAGACTCAG 8458  
 Qy 8461 GGATATCTGAGCTCTGCTCTCCACAGTGTGAATTAACAGGATACATCTGCTGTAAG 8520  
 Db 8459 GGATATCTGAGCTCTGCTCTCCACAGTGTGAATTAACAGGATACATCTGCTGTAAG 8518  
 Qy 8521 ATTTTAACTGAAATCTGAGGATAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 8580  
 Db 8519 ATTTTAACTGAAATCTGAGGATAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 8578  
 Qy 8581 GTTTGTTTGTCTCTGCTGATAGATACAGGAGCTGTAATATGTAGCTGTGGCTAC 8640  
 Db 8579 GTTTGTTTGTCTCTGCTGATAGATACAGGAGCTGTAATATGTAGCTGTGGCTAC 8638  
 Qy 8641 ATTAACATTTGTCTCAAAAAGCTATATAGAGTGAAGGAGTGTGAGGCTTAAGAGAGCTT 8700  
 Db 8639 ATTAACATTTGTCTCAAAAAGCTATATAGAGTGAAGGAGTGTGAGGCTTAAGAGAGCTT 8698  
 Qy 8701 AAGCGGCTGTGATAGCAGACAGAGTATGCTGCACTATATAGCAAGACTTGTTCAAA 8760  
 Db 8699 AAGCGGCTGTGATAGCAGACAGAGTATGCTGCACTATATAGCAAGACTTGTTCAAA 8758  
 Qy 8761 ACATGAGGAGGAGGATGTTTATGCTGTGGCTGTGTAAACAGCACTAAGAGGCCAA 8820  
 Db 8759 ACATGAGGAGGAGGATGTTTATGCTGTGGCTGTGTAAACAGCACTAAGAGGCCAA 8818  
 Qy 8821 TGTAGACATTTGACTAAGAAAGATCATCAATCAAAAGCCGGGTGGGCAAGGTGAG 8880  
 Db 8819 TGTAGACATTTGACTAAGAAAGATCATCAATCAAAAGCCGGGTGGGCAAGGTGAG 8878  
 Qy 8881 ACTACAGTGTGCAAGACCCCATATGGAAGCAGTTTCCCTTCTCTGAGGCTCAAGC 8940  
 Db 8879 ACTACAGTGTGCAAGACCCCATATGGAAGCAGTTTCCCTTCTCTGAGGCTCAAGC 8938  
 Qy 8941 CTGGCTGAGGCGCACTGTCTCAATGCTTCTCTAGGCTGTGCAACATG 8995  
 Db 8939 CTGGCTGAGGCGCACTGTCTCAATGCTTCTCTAGGCTGTGCAACATG 8993

RESULT 3  
 AAS19044  
 ID AAS19044 standard; cdna: 2628 BP.  
 XX  
 AC AAS19044;  
 XX  
 DT 09-APR-2002 (first entry)  
 XX  
 DE Mouse MU-1 haematopoietin receptor superfamily chain cdna sequence.  
 XX  
 KW Mouse; MU-1; haematopoietin receptor superfamily chain; thyroiditis;











XX	25-MAR-2003	(updated)
DT	14-JUL-1994	(first entry)
XX		
DE	Grapevine ribosomal clone INVGT19C.	
XX		
KM	Ribosome; grapevine; Vitis; IGS region; rDNA; polymorphism;	
KM	grape cultivar; probe; primer; detection; ds.	
XX		
OS	Vitis vinifera.	
XX		
FH	Key	Location/Qualifiers
FT	primer_bind	76..100
FT		/*tag= a
FT	primer_bind	complement (295..317)
FT		/*tag= b
XX		
PN	MO9401580-A1.	
XX		
PD	20-JAN-1994.	
XX		
PE	30-JUN-1993;	93WO-AU003320.
XX		
PR	03-JUL-1992;	92AU-0003330.
XX		
PA	(CSIR ) COMMONWEALTH SCI & IND RES ORG.	
XX		
PI	Scott NS, Thomas MR;	
DR	WPI; 1994-035083/04.	
XX		
PT	Novel ribosome DNA probe sequences - for the accurate identification	
PT	of grape cultivars	
XX		
PS	Claim 26; Fig 5p; 55pp; English.	
XX		
CC	The sequences given in AA055231-50 are derived from ribosomes of the	
CC	grapevine genus Vitis. These sequences represent the IGS region of	
CC	the ribosomal (r)DNA repeat and contain polymorphisms. These	
CC	polymorphisms may be used in a method for the identification of	
CC	different grape cultivars. These clones contain simple repeat	
CC	sequences and were identified in a genomic library of grapevine DNA	
CC	using simple di-, tri- or tetra- nucleotide repeats such as (AT)8,	
CC	(GT)10, (CGT)10 and such like as probes.	
CC	(Updated on 25-MAR-2003 to correct PN field.)	
XX		
SO	Sequence 382 BP; 122 A; 28 C; 95 G; 137 T; 0 other;	
	Query Match	0.7%; Score 63; DB 15; Length 382;
	Best Local Similarity	100.0%; Pred. No. 3.1e-16;
	Matches 63; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
OY	5698 TATGAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATG	5755
Db	140 TATG	199
OY	5758 TGT 5760	
Db	200 TGT 202	
	RESULT 10	
	AAA74316/C	
	AAA74316 standard; DNA; 382 BP.	
AC	AAA74316;	
XX		
DT	29-NOV-2000 (first entry)	
XX		
DE	Lobliolly pine SSR locus RIPPT1125.	
XX		
KM	Lobliolly pine; Simple Sequence Repeat; SSR; microsatellite DNA repeat;	
KM	genetic marker; mapping; inheritance study; population genetics study;	

[illegible]











GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 13, 2003, 12:10:01 ; Search time 31311 Seconds

(without alignments)  
11752.486 Million cell updates/sec

Title: US-09-877-935-1

Sequence: 1 gatctggtgaccacagagaca.....ctctagctcgtccaccatg 8995

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 2888711 seqs, 2045481386 residues

Word size : 0

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

Genembl:\*  
1: gb\_ba:\*  
2: gb\_htg:\*  
3: gb\_in:\*  
4: gb\_cm:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_sts:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vl:\*  
15: em\_ba:\*  
16: em\_fun:\*  
17: em\_hum:\*  
18: em\_in:\*  
19: em\_mu:\*  
20: em\_cm:\*  
21: em\_or:\*  
22: em\_ov:\*  
23: em\_pat:\*  
24: em\_ph:\*  
25: em\_pl:\*  
26: em\_ro:\*  
27: em\_sts:\*  
28: em\_un:\*  
29: em\_vl:\*  
30: em\_htg\_hum:\*  
31: em\_htg\_inv:\*  
32: em\_htg\_other:\*  
33: em\_htg\_mus:\*  
34: em\_htg\_pin:\*  
35: em\_htg\_rod:\*  
36: em\_htg\_mam:\*  
37: em\_htg\_vrt:\*  
38: em\_sy:\*  
39: em\_hgo\_hum:\*  
40: em\_hgo\_mus:\*  
41: em\_hgo\_other:\*

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	8995	100.0	8995	6 AX027824	AX027824 Sequence
2	4465	49.6	192060	2 AC098570	AC098570 Mus muscu
3	2086	23.2	170074	2 AC132885	AC132885 Mus muscu
4	180	2.0	779	11 BV067438	BV067438 S212P6046
5	96	1.1	149109	2 AC101018	AC101018 Rattus no
6	96	1.1	193472	2 AC1515173	AC1515173 Rattus no
7	82	0.9	145489	2 AC136730	AC136730 Mus muscu
8	82	0.9	154704	2 AC115850	AC115850 Mus muscu
9	82	0.9	173892	2 AC120393	AC120393 Mus muscu
10	82	0.9	189241	2 AC139041	AC139041 Mus muscu
11	82	0.9	190776	2 AC131335	AC131335 Mus muscu
12	82	0.9	203855	10 AC122900	AC122900 Mus muscu
13	82	0.9	222185	2 AC127697	AC127697 Mus muscu
14	82	0.9	226909	2 AC079450	AC079450 Mus muscu
15	82	0.9	232568	2 AC141471	AC141471 Mus muscu
16	82	0.9	253783	2 AC093351	AC093351 Mus muscu
17	81	0.9	181600	2 AC119189	AC119189 Mus muscu
18	81	0.9	216199	2 AC114583	AC114583 Mus muscu
19	81	0.9	247680	2 AC139943	AC139943 Mus muscu
20	80	0.9	192182	2 AC118017	AC118017 Mus muscu
21	80	0.9	235302	2 AC073784	AC073784 Mus muscu
22	80	0.9	316822	2 AC118016	AC118016 Mus muscu
23	79	0.9	57957	10 BX294655	BX294655 Mouse DNA
24	79	0.9	182061	2 AC102217	AC102217 Mus muscu
25	79	0.9	202855	10 AL589737	AL589737 Mouse DNA
26	79	0.9	223280	2 AC140409	AC140409 Mus muscu
27	78	0.9	153332	2 AC117755	AC117755 Mus muscu
28	78	0.9	204068	2 AC103395	AC103395 Mus muscu
29	78	0.9	292653	2 AC102669	AC102669 Mus muscu
30	78	0.9	316888	2 AC102670	AC102670 Mus muscu
31	77	0.9	172956	2 AC122433	AC122433 Mus muscu
32	77	0.9	182777	2 AC110034	AC110034 Mus muscu
33	76	0.8	116898	10 AL731687	AL731687 Mouse DNA
34	76	0.8	133296	10 AL935323	AL935323 Mouse DNA
35	76	0.8	178032	2 AC087869	AC087869 Mus muscu
36	76	0.8	195320	2 AC132412	AC132412 Mus muscu
37	76	0.8	199019	10 AL928812	AL928812 Mouse DNA
38	76	0.8	201602	2 AC074146	AC074146 Mus muscu
39	76	0.8	203083	2 AC069459	AC069459 Mus muscu
40	75	0.8	824	11 BV025926	BV025926 S212P6802
41	75	0.8	71848	2 AC114623_3	Continuation (4 of
42	75	0.8	131685	2 AC142273	AC142273 Mus muscu
43	75	0.8	149992	2 AC112940	AC112940 Mus muscu
44	75	0.8	168889	10 AL731779	AL731779 Mouse DNA
45	75	0.8	186858	10 AC130210	AC130210 Mus muscu

#### ALIGNMENTS

RESULT 1	AX027824	8995 bp	DNA	11near	PAT 16-SEP-2000
LOCUS	AX027824	Sequence 1 from Patent WO0034492.			
DEFINITION	AX027824				
ACCESSION	AX027824				
VERSION	AX027824.1	GI:10188668			
KEYWORDS					
SOURCE					
ORGANISM					
	Mus sp.				
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
	Mammalia; Eutheria; Rodentia; Sciurognathu; Muridae; Murinae; Mus.				
REFERENCE	1	Robine,S., Louvard,D., Pinto,D. and Jaisser,F.			
AUTHORS		Regulatory sequences of the mouse villin gene - use in transgenesis			
TITLE		Patent: WO 0034492-A 1 15-JUN-2000;			
JOURNAL					

Pred. No. is the number of results predicted by chance to have a



Db 1921 AGAGACATATGTGGCTCAGACCATCTGTAATCCAGTCCAGAGGGTCCACACCTT 1980  
QY 1981 CTTCTGGCCCTCCAGAGCACCATATAGTATACAGACATATACATGACGAGGAAACACC 2040  
Db 1981 CTTCTGGCCCTCCAGAGCACCATATAGTATACAGACATATACATGACGAGGAAACACC 2040  
QY 2041 CATACACATTAATTAATTAAGAACTTAAAGGTGATGTGTGTAACATGTGCT 2100  
Db 2041 CATACACATTAATTAATTAAGAACTTAAAGGTGATGTGTGTAACATGTGCT 2100  
QY 2101 TACACATGCTGATTGAAGACATATACAGCACACACTGAAGGGATCTGGGCTGGAG 2160  
Db 2101 TACACATGCTGATTGAAGACATATACAGCACACACTGAAGGGATCTGGGCTGGAG 2160  
QY 2161 AGATGGCTGAGGGTTAAGAGCATGCTCTCCGAAAGGAAGGCTCGAATTCAAAT 2220  
Db 2161 AGATGGCTGAGGGTTAAGAGCATGCTCTCCGAAAGGAAGGCTCGAATTCAAAT 2220  
QY 2221 CCTAGCACCATGGTGGCTCACAACCATCCATTAATGAGATCTGACACCCTTCTGCT 2280  
Db 2221 CCTAGCACCATGGTGGCTCACAACCATCCATTAATGAGATCTGACACCCTTCTGCT 2280  
QY 2281 GCATCTGAAGACAGCTGACAGCTACAGTGTACTAGATTAATTAATCTTTT 2340  
Db 2281 GCATCTGAAGACAGCTGACAGCTACAGTGTACTAGATTAATTAATCTTTT 2340  
QY 2341 TTTAAAAAATGAAGAGGGATCTGAGACCCCTCAAAAGATTAATGACGATGACG 2400  
Db 2341 TTTAAAAAATGAAGAGGGATCTGAGACCCCTCAAAAGATTAATGACGATGACG 2400  
QY 2401 GGTGATTATCTATCTGAGATTTTCTCTTCCGCTTGGCTTGGCACTGGGTGAGACAG 2460  
Db 2401 GGTGATTATCTATCTGAGATTTTCTCTTCCGCTTGGCTTGGCACTGGGTGAGACAG 2460  
QY 2461 CCCCTTTTCATTACAGAAGAGGGTCTACATTTATTCTGACAAAACAGACCTGCACT 2520  
Db 2461 CCCCTTTTCATTACAGAAGAGGGTCTACATTTATTCTGACAAAACAGACCTGCACT 2520  
QY 2521 ATGTTACTGTCTCTGACTATGAGCAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2580  
Db 2521 ATGTTACTGTCTCTGACTATGAGCAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2580  
QY 2581 ACACACACACACACACACACACACACATTCAGTCTCCAGAGCTCTTGGAGAGTCA 2640  
Db 2581 ACACACACACACACACACACACACACATTCAGTCTCCAGAGCTCTTGGAGAGTCA 2640  
QY 2641 AGAAGAGGCTGCCCTCAAAACAGATCTTCTCTTCCCTTAAAGAGACCGCATTC 2700  
Db 2641 AGAAGAGGCTGCCCTCAAAACAGATCTTCTCTTCCCTTAAAGAGACCGCATTC 2700  
QY 2701 AAGGTGGCAGAGATCTACAGGGGGCAGAGGCGAGGGGGAAGGAGGCGATGTTTC 2760  
Db 2701 AAGGTGGCAGAGATCTACAGGGGGCAGAGGCGAGGGGGAAGGAGGCGATGTTTC 2760  
QY 2761 AGAGACCTACAGCAGAGGGCAGCAGAGGATCCAGGTCAGGGCAGGGAGTGAAG 2820  
Db 2761 AGAGACCTACAGCAGAGGGCAGCAGAGGATCCAGGTCAGGGCAGGGAGTGAAG 2820  
QY 2821 CCTTCTTCCGAGGAAAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2880  
Db 2821 CCTTCTTCCGAGGAAAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2880  
QY 2881 GCTCATAAAAGTGGAGTGTGGCTCAGTCAAGAAAAGAGAAAGGAAAGGCGCTTGT 2940  
Db 2881 GCTCATAAAAGTGGAGTGTGGCTCAGTCAAGAAAAGAGAAAGGAAAGGCGCTTGT 2940  
QY 2941 GCCCACTGAGCGAGGCTCATGCTAGGAGAGATCTCAGGGGTGCCAGAGCCCCAC 3000  
Db 2941 GCCCACTGAGCGAGGCTCATGCTAGGAGAGATCTCAGGGGTGCCAGAGCCCCAC 3000  
QY 3001 CTTCTGTGCCAAGGAAACCCCAAGTGTAACTCTGGCTTGGGTCTGAGTTCCAGCTA 3060  
Db 3001 CTTCTGTGCCAAGGAAACCCCAAGTGTAACTCTGGCTTGGGTCTGAGTTCCAGCTA 3060  
Db 3061 CAAGACCCAGAGAGTCTTACTCATCCATCCATCCAGTGGCCCCCGGCGGCGGCGGCGG 3120  
QY 3061 CAAGACCCAGAGAGTCTTACTCATCCATCCATCCAGTGGCCCCCGGCGGCGGCGGCGG 3120  
Db 3121 CCCCAGACTCCCTGGCAGCTTCTAGGGCTGAGGGTGGCCAGCCCTGGTGGGGTTC 3180  
QY 3121 CCCCAGACTCCCTGGCAGCTTCTAGGGCTGAGGGTGGCCAGCCCTGGTGGGGTTC 3180  
Db 3121 CCCCAGACTCCCTGGCAGCTTCTAGGGCTGAGGGTGGCCAGCCCTGGTGGGGTTC 3180  
QY 3181 CTACCTGACAGTGAAGAGGCTCTAGGGCTGAGGGTGGCCAGCCCTGGTGGGGTTC 3240  
Db 3181 CTACCTGACAGTGAAGAGGCTCTAGGGCTGAGGGTGGCCAGCCCTGGTGGGGTTC 3240  
QY 3241 GCCAAGGGGGGCGACAGGGCAGCTCAGGCTGACGGCTGCTGGGGCTGAGGTTCC 3300  
Db 3241 GCCAAGGGGGGCGACAGGGCAGCTCAGGCTGACGGCTGCTGGGGCTGAGGTTCC 3300  
QY 3301 AGGACCTGGGCGCTACTTCCCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3360  
Db 3301 AGGACCTGGGCGCTACTTCCCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3360  
QY 3361 CTTATATGTTGAAGAAAGTCTGAGGGGGGGGGGGTGTGTGAGGACAAAGTCTTTC 3420  
Db 3361 CTTATATGTTGAAGAAAGTCTGAGGGGGGGGGGGTGTGTGAGGACAAAGTCTTTC 3420  
QY 3421 GTCTCGGACGAGCTTGGCACAACCTCTAGATCTCCAGGTGGTGGCTGCTTC 3480  
Db 3421 GTCTCGGACGAGCTTGGCACAACCTCTAGATCTCCAGGTGGTGGCTGCTTC 3480  
QY 3481 CAGACAGTAAAGCAATTTGGGTGGGACACATGGTGGACACAGGTGGTGGAGGGAG 3540  
Db 3481 CAGACAGTAAAGCAATTTGGGTGGGACACATGGTGGACACAGGTGGTGGAGGGAG 3540  
QY 3541 GGTCTTGTCTCTCTGCGACGCTGCTTCTTGTAGACACTTGGTATAGTTGGGG 3600  
Db 3541 GGTCTTGTCTCTCTGCGACGCTGCTTCTTGTAGACACTTGGTATAGTTGGGG 3600  
QY 3601 TGAAGTAAAGTGTCTGAAACCTGAAAGAGCAAGAGGACAGGCTGCTTGGGGCT 3660  
Db 3601 TGAAGTAAAGTGTCTGAAACCTGAAAGAGCAAGAGGACAGGCTGCTTGGGGCT 3660  
QY 3661 TCAATGAAGAAATTCACAGAGCCCTTCTCTGTAAGTACCTTCCATCTGATG 3720  
Db 3661 TCAATGAAGAAATTCACAGAGCCCTTCTCTGTAAGTACCTTCCATCTGATG 3720  
QY 3721 ATTCCCTGGGACCAAGTGGCTCTGGGACTGAGATTTCTACAAATTAATCGAGAG 3780  
Db 3721 ATTCCCTGGGACCAAGTGGCTCTGGGACTGAGATTTCTACAAATTAATCGAGAG 3780  
QY 3781 CCTGAGACTTGGACCTCGTGCCTGTATTACTATCTCTGCTGGCTGCTATTTCTGT 3840  
Db 3781 CCTGAGACTTGGACCTCGTGCCTGTATTACTATCTCTGCTGGCTGCTATTTCTGT 3840  
QY 3841 TCAATGCTTACACATCTGAATGTTCTTGTGTGACCAATTCCTGACACTCTGGGA 3900  
Db 3841 TCAATGCTTACACATCTGAATGTTCTTGTGTGACCAATTCCTGACACTCTGGGA 3900  
QY 3901 GGTGATATCTTGGCAGATGTATCTGGGATGTAACTGAGGCGGCGGAGGAGGAG 3960  
Db 3901 GGTGATATCTTGGCAGATGTATCTGGGATGTAACTGAGGCGGCGGAGGAGGAG 3960  
QY 3961 AGAGTGAAGAGTGTCTCTAGGCGCTATTAGGCTGAGACATACCCCTTCTAGAAAT 4020  
Db 3961 AGAGTGAAGAGTGTCTCTAGGCGCTATTAGGCTGAGACATACCCCTTCTAGAAAT 4020  
QY 4021 GGCCCTCCATTTTGGTTACCATATCATATTTATACAGAGTGGGAGGAAAGGCA 4080  
Db 4021 GGCCCTCCATTTTGGTTACCATATCATATTTATACAGAGTGGGAGGAAAGGCA 4080  
QY 4081 AACCTGCCAAGAAAGTTGGGACTCACTGAGACCAAGGTTATCTGTCAAGAAATCCCC 4140  
Db 4081 AACCTGCCAAGAAAGTTGGGACTCACTGAGACCAAGGTTATCTGTCAAGAAATCCCC 4140

```
OY 4141 TCACCTAGGTTGGGAGAAATCTGCTCTGGGGGCTTCCAGGTCTTGGTTAGCAGAGGGT 4200
|||||
Db 4141 TCACCTTAGGTTGGGAGAAATCTGCTCTGGGGGCTTCCAGGTCTTGGTTAGCAGAGGGT 4200
OY 4201 ATCCTTGTATAGGCAATGACCTAGTCTATGGTGTATCTACTACATCTGCTGCTCCAGTTAAAG 4260
|||||
Db 4201 ATCTCTGTATAGGCAATGACCTAGTCTATGGTGTATCTACTACATCTGCTGCTCCAGTTAAAG 4260
OY 4261 CTGGAAGTAAACCAGGCGAGCCGCAAGATTTCTACAGTTGTACCCCAAGAACAA 4320
|||||
Db 4261 CTGGAAGTAAACCAGGCGAGCCGCAAGATTTCTACAGTTGTACCCCAAGAACAA 4320
OY 4321 AGACAGTAGATATCCAGAGATAGTAGCTGGGAGAAAGAACTTAAACCCCAAG 4380
|||||
Db 4321 AGACAGTAGATATCCAGAGATAGTAGCTGGGAGAAAGAACTTAAACCCCAAG 4380
OY 4381 GCCCAGAGTTCCGCTTCCCTAGTTCACAATGCCAGTGTAGTGTAGTACTATGGGCTG 4440
|||||
Db 4381 GCCCAGAGTTCCGCTTCCCTAGTTCACAATGCCAGTGTAGTGTAGTACTATGGGCTG 4440
OY 4441 TGAGTTGGTAGCTCAAGCATGATGATGTTGATGTGTAGTGTATATATCTGAGCAC 4500
|||||
Db 4441 TGAGTTGGTAGCTCAAGCATGATGATGTTGATGTGTAGTGTATATATCTGAGCAC 4500
OY 4501 TTGGAGGCTGAGCAGAGAGATTGCTATATGTTTGAAGGCCAGCCTGAGCTATAGAGCA 4560
|||||
Db 4501 TTGGAGGCTGAGCAGAGAGATTGCTATATGTTTGAAGGCCAGCCTGAGCTATAGAGCA 4560
OY 4561 GACCTTTCCTTAAAGAAAAATGAAAGCCAGAGGTGGGACAGGCCCTTTATGCCA 4620
|||||
Db 4561 GACCTTTCCTTAAAGAAAAATGAAAGCCAGAGGTGGGACAGGCCCTTTATGCCA 4620
OY 4621 GCACCTTGGAGGCAAGAGCAGAGATTTCTGAGTTCAAGGCCAGCCTGCTCTATAGAGT 4680
|||||
Db 4621 GCACCTTGGAGGCAAGAGCAGAGATTTCTGAGTTCAAGGCCAGCCTGCTCTATAGAGT 4680
OY 4681 GAGTTTCCAGCAGACCAGGGCTACACAGAGAAACCTGTTTGAAGAAAAACCGAAGAAAA 4740
|||||
Db 4681 GAGTTTCCAGCAGACCAGGGCTACACAGAGAAACCTGTTTGAAGAAAAACCGAAGAAAA 4740
OY 4741 AACAAAAAAGAAAAAAGAAAAACCAACCCAAACCCAAACCTCATCTCTCATCTCTC 4800
|||||
Db 4741 AACAAAAAAGAAAAAAGAAAAACCAACCCAAACCCAAACCTCATCTCTCATCTCTC 4800
OY 4801 TAGGCTGTGTCTCTAGTGTGTAGAGTTTGGGGACTTCAGACTTATATATATATAGGCC 4860
|||||
Db 4801 TAGGCTGTGTCTCTAGTGTGTAGAGTTTGGGGACTTCAGACTTATATATATATAGGCC 4860
OY 4861 TTTTATACACTGCTCAGAGACGAGAAAGTTTCACTGTGGGACACAGTGGGACCTTGAGA 4920
|||||
Db 4861 TTTTATACACTGCTCAGAGACGAGAAAGTTTCACTGTGGGACACAGTGGGACCTTGAGA 4920
OY 4921 AAGTACCTCTGGCAGGCCAAAAATTTCTGGGAAGGCTTCTGGGGAAGTGTGCCGAT 4980
|||||
Db 4921 AAGTACCTCTGGCAGGCCAAAAATTTCTGGGAAGGCTTCTGGGGAAGTGTGCCGAT 4980
OY 4981 CAGACTACTGTTCTTGAAGGCAAGAGAGGTTGGAAGATTTTGTGTGACAGACAGTT 5040
|||||
Db 4981 CAGACTACTGTTCTTGAAGGCAAGAGAGGTTGGAAGATTTTGTGTGACAGACAGTT 5040
OY 5041 GGAACAGAAAGACAGAGGGGAGGCAATCCAAAGATTCTGAACATGTAGCTACTTTTGGT 5100
|||||
Db 5041 GGAACAGAAAGACAGAGGGGAGGCAATCCAAAGATTCTGAACATGTAGCTACTTTTGGT 5100
OY 5101 TCTTGGGTGACAAGTGTCCCCCAAGGATAGGGCTGTAGAAAGGGGACAGGGGTGAGCC 5160
|||||
Db 5101 TCTTGGGTGACAAGTGTCCCCCAAGGATAGGGCTGTAGAAAGGGGACAGGGGTGAGCC 5160
OY 5161 AATGAGTTCAAGTTGAGGAGACATCCAGGCCAGGGCTTGTGCTGGCAAGTTAAAGATG 5220
|||||
Db 5161 AATGAGTTCAAGTTGAGGAGACATCCAGGCCAGGGCTTGTGCTGGCAAGTTAAAGATG 5220

OY 5221 AGACCCCTCTAACCCCTCCCTGAAGTTTAAAGGGAGACAGAGAGCTGAGAGATCTTCTA 5280
|||||
Db 5221 AGACCCCTCTAACCCCTCCCTGAAGTTTAAAGGGAGACAGAGAGCTGAGAGATCTTCTA 5280
OY 5281 GGGTGAAGAGAGGATATCTGCTGACCAACATGGCTAGAGACAGAGCATGTTGGACAG 5340
|||||
Db 5281 GGGTGAAGAGAGGATATCTGCTGACCAACATGGCTAGAGACAGAGCATGTTGGACAG 5340
OY 5341 TTACCCCTCAGAACACAGCATCCCTCTTGGCTCTAAGAGGCTGGGCCCTTCTGTCT 5400
|||||
Db 5341 TTACCCCTCAGAACACAGCATCCCTCTTGGCTCTAAGAGGCTGGGCCCTTCTGTCT 5400
OY 5401 AAGAACTTACTTTCTTCAAGAGAGGAGCAGACCTTTGTCCCTCCCTGTGGTCA 5460
|||||
Db 5401 AAGAACTTACTTTCTTCAAGAGAGGAGCAGACCTTTGTCCCTCCCTGTGGTCA 5460
OY 5461 TAAACACCCCTGTGTGAACATTTGTTATTTTACTGTCAAGTTGTCTCCAGAGATCCA 5520
|||||
Db 5461 TAAACACCCCTGTGTGAACATTTGTTATTTTACTGTCAAGTTGTCTCCAGAGATCCA 5520
OY 5521 TCTGTAGACCTCTGCTCTTAACCTCACCAGATGAGCCCAATYCTCACCAGAAAG 5580
|||||
Db 5521 TCTGTAGACCTCTGCTCTTAACCTCACCAGATGAGCCCAATYCTCACCAGAAAG 5580
OY 5581 TGCAAGAGAGGCTTAGAGAAAGGATACAGTAAACAAATGGCCAGAAATTAACAAA 5640
|||||
Db 5581 TGCAAGAGAGGCTTAGAGAAAGGATACAGTAAACAAATGGCCAGAAATTAACAAA 5640
OY 5641 ACTACTATCCCTTTTACCACAAATGTTTGTCTGAACAGAGGGGCTGTGTAT 5700
|||||
Db 5641 ACTACTATCCCTTTTACCACAAATGTTTGTCTGAACAGAGGGGCTGTGTAT 5700
OY 5701 GTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 5760
|||||
Db 5701 GTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 5760
OY 5761 CTTGGGGGACTTTTCAATCCTTAAAGATATCGATATGGGGCCATGCCAACAGGGAT 5820
|||||
Db 5761 CTTGGGGGACTTTTCAATCCTTAAAGATATCGATATGGGGCCATGCCAACAGGGAT 5820
OY 5821 TGGGAGAGTACAGCTTCTGCAAAACACAGTAAAGTGGCCCAAGATGATTTGGTGGCTGAA 5880
|||||
Db 5821 TGGGAGAGTACAGCTTCTGCAAAACACAGTAAAGTGGCCCAAGATGATTTGGTGGCTGAA 5880
OY 5881 TCACCAAGGGGAGGCTGATCAGATGAGACAGAAATCACAAGATTAAGCCACCTGTGGG 5940
|||||
Db 5881 TCACCAAGGGGAGGCTGATCAGATGAGACAGAAATCACAAGATTAAGCCACCTGTGGG 5940
OY 5941 GCTCAGAAAGGGGAGTTTACAAGAGGTTAAAGGCCCAAGCTTTTATCCAAAGCATGAC 6000
|||||
Db 5941 GCTCAGAAAGGGGAGTTTACAAGAGGTTAAAGGCCCAAGCTTTTATCCAAAGCATGAC 6000
OY 6001 TCAAAATCAAAAGTCAAGAGAGATTTAGCTGTGAGAGATGGGGCTGTCAAGTGTGGACACC 6060
|||||
Db 6001 TCAAAATCAAAAGTCAAGAGAGATTTAGCTGTGAGAGATGGGGCTGTCAAGTGTGGACACC 6060
OY 6061 TGACCTTGACTTATATGTCATAGGCCAAGGACAGTACAGAGGGTGAAGTGGCTCTTA 6120
|||||
Db 6061 TGACCTTGACTTATATGTCATAGGCCAAGGACAGTACAGAGGGTGAAGTGGCTCTTA 6120
OY 6121 CTGAGCTTGGAGGAGGACGCGTGGAGATGGGTGACGTCACATCCAGATGAGAGGGCTGAG 6180
|||||
Db 6121 CTGAGCTTGGAGGAGGACGCGTGGAGATGGGTGACGTCACATCCAGATGAGAGGGCTGAG 6180
OY 6181 CACCACCAAGGTACAAGTGTCCCTGTGTCTCATCCAGATTTCTTGCCAGATTTTCAAG 6240
|||||
Db 6181 CACCACCAAGGTACAAGTGTCCCTGTGTCTCATCCAGATTTCTTGCCAGATTTTCAAG 6240
OY 6241 GACTTAAGACTCATCTGTGTGGAACAAAGTATCCAAAGCCCTTAAGCCCTATTTTGGTCT 6300
|||||
Db 6241 GACTTAAGACTCATCTGTGTGGAACAAAGTATCCAAAGCCCTTAAGCCCTATTTTGGTCT 6300
OY 6301 AATTAATTCAGAACCCCTGGGAGTGCAGGCTCTGACAGCAGAGGCTTTTAAAGACCTC 6360
|||||
```

|||||  
Db 6301 AATTAAATCAGAACCCCTGGGGATGACAGGCTCTAGACGACGAGGAGCTTTTAAAAAGCTC 6360  
Qy 6361 CCAGGATTTCTGATTCAGACGCTGGAAACAAACACAGCTACAGGTTCAAACGAAGAGGC 6420  
Db 6361 CCAGGATTTCTGATTCAGACGCTGGAAACAAACACAGCTACAGGTTCAAACGAAGAGGC 6420  
Qy 6421 AAAGCTAGGGAACCTTGGATGGGAGGACCTTCTTCCAGGCCAGTAGATGGAGGCTGTT 6480  
Db 6421 AAAGCTAGGGAACCTTGGATGGGAGGACCTTCTTCCAGGCCAGTAGATGGAGGCTGTT 6480  
Qy 6481 AGCAGTGGTGGCAGCTTCTCTGCTGTCTATATAGCTATTCATCCATCCATCCATCAT 6540  
Db 6481 AGCAGTGGTGGCAGCTTCTCTGCTGTCTATATAGCTATTCATCCATCCATCCATCAT 6540  
Qy 6541 ACACCCAGCCATCATTTATGACACCCATCCCTCCATCCATCCATCCATCCATCCATCC 6600  
Db 6541 ACACCCAGCCATCATTTATGACACCCATCCCTCCATCCATCCATCCATCCATCCATCC 6600  
Qy 6601 CCACGATCCATCCAAACCTTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 6660  
Db 6601 CCACGATCCATCCAAACCTTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 6660  
Qy 6661 CATTTATCCAAAGAGACTGCTGTAATGCTAAATGCTGAGATTTAATTTTGA 6720  
Db 6661 CATTTATCCAAAGAGACTGCTGTAATGCTAAATGCTGAGATTTAATTTTGA 6720  
Qy 6721 AGGCTGTGATGACGATTTGACATGATGACATGACATGACATGACATGACATGACATG 6780  
Db 6721 AGCTCTGTGATGACGATTTGACATGATGACATGACATGACATGACATGACATGACATG 6780  
Qy 6781 TGGCAATCGGAGAAAGGTTTGGGTGTGTTTCTCTTCCACGCTGTGGGTCTGGGA 6840  
Db 6781 TGGCAATCGGAGAAAGGTTTGGGTGTGTTTCTCTTCCACGCTGTGGGTCTGGGA 6840  
Qy 6841 TTGACATCAATTAATCGGGCTGTGGCAAGTGTCTTTACACGAGCATTTTCTGACA 6900  
Db 6841 TTGACATCAATTAATCGGGCTGTGGCAAGTGTCTTTACACGAGCATTTTCTGACA 6900  
Qy 6901 CATCATTAATTTGAAGAGCATTTATGATGATGATGATGATGATGATGATGATGATGATG 6960  
Db 6901 CATCATTAATTTGAAGAGCATTTATGATGATGATGATGATGATGATGATGATGATGATG 6960  
Qy 6961 CAAGGATGACCTTTAATCTGCTCTCTTCCAGCTCCACCGAGTGTAGGTTTACAGGT 7020  
Db 6961 CAAGGATGACCTTTAATCTGCTCTCTTCCAGCTCCACCGAGTGTAGGTTTACAGGT 7020  
Qy 7021 GTTCACTGGTGAATGCTTTAATCCAGCAGCTGTGGGGGGGGGGGGGGGGGGGGGGGGG 7080  
Db 7021 GTTCACTGGTGAATGCTTTAATCCAGCAGCTGTGGGGGGGGGGGGGGGGGGGGGGGGG 7080  
Qy 7081 CTTGAGTGGAGGCTGTTGGTGTACAGAGTTCAAGATTAACCTGGGGCTTATACAGGAA 7140  
Db 7081 CTTGAGTGGAGGCTGTTGGTGTACAGAGTTCAAGATTAACCTGGGGCTTATACAGGAA 7140  
Qy 7141 ACCCTATCCAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAA 7200  
Db 7141 ACCCTATCCAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAA 7200  
Qy 7201 TTAGAGATTTAGTAGGGTAGTAGGGCTGGTGAGGAGATCATGCTTTCTTTTGAAT 7260  
Db 7201 TTAGAGATTTAGTAGGGTAGTAGGGCTGGTGAGGAGATCATGCTTTCTTTTGAAT 7260  
Qy 7261 AATATAGTAAGTACTACAAAGATGATTAATCTATCTATCTATCTATCTATCTATCTATCT 7320  
Db 7261 AATATAGTAAGTACTACAAAGATGATTAATCTATCTATCTATCTATCTATCTATCTATCT 7320  
Qy 7321 TATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTAT 7380  
Db 7321 TATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTAT 7380  
Qy 7381 CTGCTTGAAGTCTGAATGCTCTATTTCTGGGTCACTCTTACCCCTAGTGTGGGT 7440  
Db 7381 CTGCTTGAAGTCTGAATGCTCTATTTCTGGGTCACTCTTACCCCTAGTGTGGGT 7440

|||||  
Db 7441 ACCAAGCCAGACATTTATTTTGTGTTTATTTATTTATTTATTTATTTATTTATTTATTT 7500  
Qy 7441 ACCAAGCCAGACATTTATTTTGTGTTTATTTATTTATTTATTTATTTATTTATTTATTT 7500  
Db 7441 ACCAAGCCAGACATTTATTTTGTGTTTATTTATTTATTTATTTATTTATTTATTTATTT 7500  
Qy 7501 GGAAGTGGGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 7560  
Db 7501 GGAAGTGGGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 7560  
Qy 7561 CCATTTGTTAGGTAAGTCTGTGACAGTTGCTATATTTGACGAGCTATGAGTCTCTC 7620  
Db 7561 CCATTTGTTAGGTAAGTCTGTGACAGTTGCTATATTTGACGAGCTATGAGTCTCTC 7620  
Qy 7621 CACCTCCAGTTCCAGCACTTTCGTCATCCGATGCGGGGGGGGGGGGGGGGGGGGGGGGG 7680  
Db 7621 CACCTCCAGTTCCAGCACTTTCGTCATCCGATGCGGGGGGGGGGGGGGGGGGGGGGGGG 7680  
Qy 7681 GTGCTGTGTTCCCTGTCTGACAGCTTACATATTTGCGTGTGAAAGTTGATTAATG 7740  
Db 7681 GTGCTGTGTTCCCTGTCTGACAGCTTACATATTTGCGTGTGAAAGTTGATTAATG 7740  
Qy 7741 GGATGCTTCTGTGATATCTTTTATGAGCTGGCCCTTTATCTTATGACAGCTTGTGTG 7800  
Db 7741 GGATGCTTCTGTGATATCTTTTATGAGCTGGCCCTTTATCTTATGACAGCTTGTGTG 7800  
Qy 7801 GGCAATGTCACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 7860  
Db 7801 GGCAATGTCACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 7860  
Qy 7861 TGTGTGATTAACACACTTCTGTTTCAATTAATGAAATTTGTGGCCACCCAC 7920  
Db 7861 TGTGTGATTAACACACTTCTGTTTCAATTAATGAAATTTGTGGCCACCCAC 7920  
Qy 7921 CTTTTTTTTTTTTTTTATGAGCAAGGCTTCTGTGATGATGATGATGATGATGATGATG 7980  
Db 7921 CTTTTTTTTTTTTTTTATGAGCAAGGCTTCTGTGATGATGATGATGATGATGATGATG 7980  
Qy 7981 GAGCTCACTGTGTGACAGGCTGTGAGGCTGTGAGGCTGTGAGGCTGTGAGGCTGTGAGG 8040  
Db 7981 GAGCTCACTGTGTGACAGGCTGTGAGGCTGTGAGGCTGTGAGGCTGTGAGGCTGTGAGG 8040  
Qy 8041 AGAGTACCATGAACTTCAAGCAATTTCTGTTTGGTTTGGTTTGGTTTGGTTTGGTTTGG 8100  
Db 8041 AGAGTACCATGAACTTCAAGCAATTTCTGTTTGGTTTGGTTTGGTTTGGTTTGGTTTGG 8100  
Qy 8101 TGTATGCTGTATATGTCATGTTGTGTCTTCAAGGCTGTACATGTCATGTCATGTCATG 8160  
Db 8101 TGTATGCTGTATATGTCATGTTGTGTCTTCAAGGCTGTACATGTCATGTCATGTCATG 8160  
Qy 8161 TGGGACAGAGAAACCAACCGATGTGCAATCTCTGATGATGATGATGATGATGATGATG 8220  
Db 8161 TGGGACAGAGAAACCAACCGATGTGCAATCTCTGATGATGATGATGATGATGATGATG 8220  
Qy 8221 TGTATTAATGATGTTTATTTAGTGTGCCAAGTATGACGATTTTGTGGAGTTTTCAC 8280  
Db 8221 TGTATTAATGATGTTTATTTAGTGTGCCAAGTATGACGATTTTGTGGAGTTTTCAC 8280  
Qy 8281 CTTCCCTTGTGGCTGTCCGATTAATCACTGCTCTGCGGCTGTAGTGTAGTGTAGTGTAG 8340  
Db 8281 CTTCCCTTGTGGCTGTCCGATTAATCACTGCTCTGCGGCTGTAGTGTAGTGTAGTGTAG 8340  
Qy 8341 CTGATGAGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 8400  
Db 8341 CTGATGAGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 8400  
Qy 8401 CGCAGTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 8460  
Db 8401 CGCAGTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 8460  
Qy 8461 GGATATGCTGGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCT 8520  
Db 8461 GGATATGCTGGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCT 8520

```

Oy 8521 ATTTTAACTGAAATCTGAGATAGACAGGACCTACCAATGAGGTTCTTTTGT 8580
Db 8521 ATTTTAACTGAAATCTGAGATAGACAGGACCTACCAATGAGGTTCTTTTGT 8580
Oy 8581 GTTTGGTTGGTTCTCTGATAGATAGACAGGACCTGGAATAGTATGAGCTGGGCTAC 8640
Db 8581 GTTTGGTTGGTTCTCTGATAGATAGACAGGACCTGGAATAGTATGAGCTGGGCTAC 8640
Oy 8641 ATACATCTGTCTCCTCAAAAGCCTATAGAGTAGAGGAGGAGGCTTAAGAGAGCCTT 8700
Db 8641 ATACATCTGTCTCCTCAAAAGCCTATAGAGTAGAGGAGGAGGCTTAAGAGAGCCTT 8700
Oy 8701 AAGCGGCTGTGATAGACACAGATAGCCTGACATATATAGCAAGACCTTGTTCAAA 8760
Db 8701 AAGCGGCTGTGATAGACACAGATAGCCTGACATATATAGCAAGACCTTGTTCAAA 8760
Oy 8761 ACATGAGAGGAGGGGTATGTTTAAAGTCTGGGCTGTATACAGGACCTAAGGAGCCAA 8820
Db 8761 ACATGAGAGGAGGGGTATGTTTAAAGTCTGGGCTGTATACAGGACCTAAGGAGCCAA 8820
Oy 8821 TGTAGACATTTGACTAAGAAAGATCATCATCAAAAGCCGGGTGGGAGGTTAGGTTGG 8880
Db 8821 TGTAGACATTTGACTAAGAAAGATCATCATCAAAAGCCGGGTGGGAGGTTAGGTTGG 8880
Oy 8881 ACTACAGTGTCAAGACCCCATAGAGAGCCAGTTTCCCTTCTCTGGGCTCAAGC 8940
Db 8881 ACTACAGTGTCAAGACCCCATAGAGAGCCAGTTTCCCTTCTCTGGGCTCAAGC 8940
Oy 8941 CTGGCTCAGCGCCACTCTCTCACATGCTTCTCTAGGCTGTGCACCATG 8995
Db 8941 CTGGCTCAGCGCCACTCTCTCACATGCTTCTCTAGGCTGTGCACCATG 8995

```

```

RESULT 2
LOCUS AC098570 192060 bp DNA linear HTG 20-AUG-2002
DEFINITION Mus musculus clone RP23-278N11, WORKING DRAFT SEQUENCE, 6 unordered
ACCESSION AC098570.2 GI:22325297
VERSION AC098570.2
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULFILLOR.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 192060)
Birren, B., Nusbaum, C. and Lander, E.
Mus musculus, clone RP23-278N11
Unpublished
2 (bases 1 to 192060)
Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,
Anderson, S., Barna, N., Bastien, V., Boguslavsky, L., Boukhalter, B.,
Brown, A., Camarata, J., Campolano, A., Chang, J., Chazaro, B.,
Chepel, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A.,
Cooke, P., Deatellano, K., Dewar, K., Diaz, J.S., Dodge, S., Fato, S.,
Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gaidyna, S.,
Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,
Hagos, B., Heaford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R.,
Jones, C., Kamel, A., Karatas, A., Kells, C., LaRoque, K.,
Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Liu, G.,
Maclean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C.,
McCarthy, M., McEwan, P., McKernan, K., McPheters, R., Meldrum, J.,
Meneus, L., Mihova, T., Mienga, V., Murphy, T., Naylor, J., Nauyen, C.,
Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D.,
Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V.,
Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P.,
Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupback, R.,
Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
Strauss, N., Subramanian, A., Talamas, J., Testfaye, S., Theodore, J.,
Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H.,
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G.,
Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

```

TITLE  
JOURNAL  
REFERENCE  
AUTHORS

TITLE  
JOURNAL  
COMMENT

Direct Submission  
Submitted (24-Oct-2001) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
3 (bases 1 to 192060)  
Birren, B., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S.,  
Barna, N., Bastien, V., Bloom, T., Boguslavsky, L., Boukhalter, B.,  
Camarata, J., Chang, J., Chazaro, B., Chepel, Y., Collymore, A.,  
Cook, A., Cooke, P., Deatellano, K., Dewar, K., Diaz, J.S., Dodge, S.,  
Fato, S., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J.,  
Gaidyna, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,  
Hagos, B., Heaford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R.,  
Jones, C., Kamel, A., Karatas, A., Kells, C., LaRoque, K.,  
Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Liu, G.,  
Maclean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C.,  
McCarthy, M., McEwan, P., McKernan, K., McPheters, R., Meldrum, J.,  
Meneus, L., Mihova, T., Mienga, V., Murphy, T., Naylor, J., Nauyen, C.,  
Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D.,  
Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V.,  
Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P.,  
Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupback, R.,  
Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,  
Strauss, N., Subramanian, A., Talamas, J., Testfaye, S., Theodore, J.,  
Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H.,  
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J.,  
Zembek, L., Zimmer, A. and Zody, M.

Direct Submission  
Submitted (20-AUG-2002) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Aug 20, 2002 this sequence version replaced gi:16356898.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR  
Web site: <http://www-seq.wi.mit.edu>  
Contact: [sequence.submissions@genome.wi.mit.edu](mailto:sequence.submissions@genome.wi.mit.edu)

Project Information  
Center project name: 278\_N\_11  
Center clone name: 278\_N\_11

Summary Statistics  
Sequencing vector: Plasmid; n/a; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.960721  
Consensus quality: 190464 bases at least Q40  
Consensus quality: 191061 bases at least Q40  
Consensus quality: 191345 bases at least Q20  
Insert size: 163000; agarose-fp  
Insert size: 191560; sum-of-ctrls  
Quality coverage: 12.4 in Q20 bases; agarose-fp  
Quality coverage: 10.5 in Q20 bases; sum-of-ctrls

NOTE: This is a 'working draft' sequence. It currently  
consists of 6 contigs. The true order of the pieces  
is not known and their order in this sequence record is  
arbitrary. Gaps between the contigs are represented as  
runs of N, but the exact sizes of the gaps are unknown.  
This record will be updated with the finished sequence  
as soon as it is available and the accession number will  
be preserved.

1 4921: contig of 4921 bp in length  
\* 4922 5021: gap of 100 bp  
\* 5022 15200: contig of 10179 bp in length  
\* 15201 15300: gap of 100 bp  
\* 15301 78692: contig of 63392 bp in length  
\* 78693 78792: gap of 100 bp  
\* 78793 116118: contig of 37326 bp in length  
\* 116119 116218: gap of 100 bp  
\* 116219 174491: contig of 58273 bp in length  
\* 174492 174591: gap of 100 bp  
\* 174592 192060: contig of 17469 bp in length.

Location/Qualifiers  
1. 192060  
/organism="Mus musculus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10090"

FEATURES  
source



[illegible]

QY	721	GCACCTTACACATACACGCTGTCAATATGCTACACTCCCTACTGTCCACACATGCGTGTAC	780
Db	40881	GCACCTTACACATACACGCTGTCAATATGCTACACTCCCTACTGTCCACACATGCGTGTAC	40944
QY	781	CTCGTTTGTGCTTTCCATCTGTGTTGACATTTGTGCAGACCAGAAATTAGAATGTGGGT	840
Db	40941	CTCGTTTGTGCTTTCCATCTGTGTTGACATTTGTGCAGACCAGAAATTAGAATGTGGGT	41000
QY	841	ATTATTTGTGTGCGAGAGACACATCCAGGGCTTTTACATTCAGGACCATGTTTTAC	900
Db	41001	ATTATTTGTGTGCGAGAGACACATCCAGGGCTTTTACATTCAGGACCATGTTTTAC	41060
QY	901	TAACTGGGCTACTTCTCCACAGGTTTGAACCATTTGTTTATATTTACTATTTTGTGT	960
Db	41061	TAACTGGGCTACTTCTCCACAGGTTTGAACCATTTGTTTATATTTACTATTTTGTGT	41122
QY	961	GCATGAGGTAGGCAATGTATACGTATGTATAGGAGTCAATGATGTGGCTGTACCCCTCAA	1020
Db	41121	GCATGAGGTAGGCAATGTATACGTATGTATAGGAGTCAATGATGTGGCTGTACCCCTCAA	41188
QY	1021	ATCATTTGAGATTTCCACAGTGAAGTACCGAGCGTTTGAATGTTATATGTGGAGTGT	1080
Db	41181	ATCATTTGAGATTTCCACAGTGAAGTACCGAGCGTTTGAATGTTATATGTGGAGTGT	41244
QY	1081	GGAGCCAGAGCTGGGGTCTCTGCAAGACACCAAGTGGCTTTAAACAATGGGACAGCTCT	1140
Db	41241	GGAGCCAGAGCTGGGGTCTCTGCAAGACACCAAGTGGCTTTAAACAATGGGACAGCTCT	41300
QY	1141	CTAGGCGTAAAGTAACTTTAGTTTTTAAAAATATATATTTCTACAGCCGGGTGTGTGGC	1200
Db	41301	CTAGGCGTAAAGTAACTTTAGTTTTTAAAAATATATATTTCTACAGCCGGGTGTGTGGC	41366
QY	1201	ACAGCGCTTTATCCACACACTTAGAGCGGAGGTGTAGGAATTTATACACAGGCGAG	1260
Db	41361	ACAGCGCTTTATCCACACACTTAGAGCGGAGGTGTAGGAATTTATACACAGGCGAG	41422
QY	1261	CTGGGGTCAAGACTTGCGCCCTGTTTTTTTTGTTTTTCTTATATGTGCACGTGTCTTA	1320
Db	41421	CTGGGGTCAAGACTTGCGCCCTGTTTTTTTTGTTTTTCTTATATGTGCACGTGTCTTA	41488
QY	1321	CTCGCGTATGTCCGTGCAAGGGGTGCAGATCCCTTGGAGCTGGAGTTAAAGACAGTTG	1380
Db	41481	CTCGCGTATGTCCGTGCAAGGGGTGCAGATCCCTTGGAGCTGGAGTTAAAGACAGTTG	41544
QY	1381	TGATACAGCGTCCGTTACAGATGCTGGGAAATTTGAACCCAGGTTGCCATAGGAAGACGCC	1440
Db	41541	TGATACAGCGTCCGTTACAGATGCTGGGAAATTTGAACCCAGGTTGCCATAGGAAGACGCC	41600
QY	1441	AGTGTCTTAATCTTGTGAGCACCCCTTCAACCCCTGTTTATAGAGACTCTTAACCTTTTG	1500
Db	41601	AGTGTCTTAATCTTGTGAGCACCCCTTCAACCCCTGTTTATAGAGACTCTTAACCTTTTG	41666
QY	1501	TGTATATGTGGAACTGAGTGGATCTTGCATTTACCAAGTGTGTGCTGCGCTGTAGCATCA	1560
Db	41661	TGTATATGTGGAACTGAGTGGATCTTGCATTTACCAAGTGTGTGCTGCGCTGTAGCATCA	41722
QY	1561	CTGAGCCCGTACCCACACAGACTAGTGGATTAAGTTTAAAGGCAAAACCTTAACANTAGA	1620
Db	41721	CTGAGCCCGTACCCACACAGACTAGTGGATTAAGTTTAAAGGCAAAACCTTAACANTAGA	41788
QY	1621	ATTAGTTGATAGATTTGAATATATAGTCCGTGAGCTATTGTTAGCGTGACCTTCTGTCC	1680
Db	41781	ATTAGTTGATAGATTTGAATATATAGTCCGTGAGCTATTGTTAGCGTGACCTTCTGTCC	41844
QY	1681	TTAGCATGTGCTGTGAGAGATTAAGAAAAATGAGAACTTGAAGTCTAAGTCTTGAACCCACA	1740
Db	41841	TTAGCATGTGCTGTGAGAGATTAAGAAAAATGAGAACTTGAAGTCTAAGTCTTGAACCCACA	41900
QY	1741	GAGGCAGGCGAGAACCCACATCTCCTAAAGTTGTCTGTAGCTTACATTAACAATTCACAT	1800
Db	41901	GAGGCAGGCGAGAACCCACATCTCCTAAAGTTGTCTGTAGCTTACATTAACAATTCACAT	41966
QY	1801	AATAGTTACATATGATTAATTAATTAATTTAGTAAATTTCTTTTAAAGGTATATATGTGGAGGGA	1860



D	b		41961	AATGTTACAAATGTATAATTAATTAGTAAATCTTTTAGAAGGTATATGTGGAGGGA	42020
Oy			1861	GAGATGGCTCAGCTTCAGAGACACTTGCTCTTCAGAGACCCTAGATTCCAGTTCCC	1920
D	b		42021	GAGATGGCTCAGCTTCAGAGACACTTGCTCTTCAGAGACCCTAGATTCCAGTTCCC	42080
Oy			1921	AGGACTCATATGTGTGGCTCACAGCACTCTGTAAATCCAGTTCCAGAGGTTCCACACCTT	1980
D	b		42081	AGGACTCATATGTGTGGCTCACAGCACTCTGTAAATCCAGTTCCAGAGGTTCCACACCTT	42140
Oy			1981	CTTGCGGCCTCCACAGGACACCATATAGTACAGACATATCATGACAGCAAAACACC	2040
D	b		42141	CTTGCGGCCTCCACAGGACACCATATAGTACAGACATATCATGACAGCAAAACACC	42200
Oy			2041	CATACACACATAAATTAATTAAGAACAATTAAGAAGTGCATGTGTGTAAACATTTGCT	2100
D	b		42201	CATACACACATAAATTAATTAAGAACAATTAAGAAGTGCATGTGTGTAAACATTTGCT	42260
Oy			2101	TACACATGCTGATTTGAAGACATGTACAAAGCAGACACTGAAGAGGATCTGGGCTGGAG	2160
D	b		42261	TACACATGCTGATTTGAAGACATGTGTACAAAGCAGACACTGAAGAGGATCTGGGCTGGAG	42320
Oy			2161	AGATGGCTCAGCGGTTAAGAGCACTGACTGCTCTCCGAAGGAAGTCTGAGTTCAAT	2220
D	b		42321	AGATGGCTCAGCGGTTAAGAGCACTGACTGCTCTCTCCGAAGGAAGTCTGAGTTCAAT	42380
Oy			2221	CCTAGCAACACATGTGGGCTCACACACCATCATATAGATGTGACACCCCTCTTCGGT	2280
D	b		42381	CCTAGCAACACATGTGGGCTCACACACCATCATATAGATGTGACACCCCTCTTCGGT	42440
Oy			2281	GCATCTGAAGACAGCTGCAGACGTACTAGTACTATATATATTAATTAATCTTTT	2340
D	b		42441	GCATCTGAAGACAGCTGCAGACGTACTAGTACTATATATATTAATTAATCTTTT	42500
Oy			2341	TTTTAAAAAATGAAGAGGATCTGAGACACCTCAAAAAGATTTGTGACAGTGCACG	2400
D	b		42501	TTTTAAAAAATGAAGAGGATCTGAGACACCTCAAAAAGATTTGTGACAGTGCACG	42560
Oy			2401	GGTGAATATCTATCTGTGAGTTTTTCCCTTCCGCTTGCAACTGGGTGGACAGACG	2460
D	b		42561	GGTGAATATCTATCTGTGAGTTTTTCCCTTCCGCTTGCAACTGGGTGGACAGACG	42620
Oy			2461	CCCCCTTTTCATTCACAGAACGGGTGCTACATTTATTTCTGAACAAAACAGCACTGCAGT	2520
D	b		42621	CCCCCTTTTCATTCACAGAACGGGTGCTACATTTATTTCTGAACAAAACAGCACTGCAGT	42680
Oy			2521	ATGTTTACTGTGCTGTGACATATAGAGAGCGGAGCGGCGGCGACACACACACAC	2578
D	b		42681	ATGTTTACTGTGCTGTGACATATAGAGAGCGGAGCGGCGGCGACACACACACAC	42740
Oy			2579	ACACACACACACACACACACACACACACATTCAGTCTCCAGAGCTTTGGGAAGT	2638
D	b		42741	ACACACACACACACACACACACACACACATTCAGTCTCCAGAGCTTTGGGAAGT	42800
Oy			2639	CAAGAAAGAGGTGCCCTCTCAAACAGATCTTCATCTTCCTCTTAAGAGAGACACGAT	2698
D	b		42801	CAAGAAAGAGGTGCCCTCTCAAACAGATCTTCATCTTCCTCTTAAGAGAGACACGAT	42860
Oy			2699	CCAAGGTGCGAAGATCTTACAGGGGCGAGAGGAGGAGGAGGAGGAGGAGGAGGAGT	2758
D	b		42861	CCAAGGTGCGAAGATCTTACAGGGGCGAGAGGAGGAGGAGGAGGAGGAGGAGGAGT	42920
Oy			2759	CCAGAGACCTTCAGAGAGGCGCAGCAAGGAGATCCAGAGTCCAGGCGAGAGAGTGA	2818
D	b		42921	CCAGAGACCTTCAGAGAGGCGCAGCAAGGAGATCCAGAGTCCAGGCGAGAGAGTGA	42980
Oy			2819	GGCCCTTGTCCGAGAGAAAGCGAGCGGCGCAAGAGAGGTTCAAGAGCACAGTTTATGG	2878
D	b		42981	GGCCCTTGTCCGAGAGAAAGCGAGCGGCGCAAGAGAGGTTCAAGAGCACAGTTTATGG	43040
Oy			2879	CAGCTCATAAAGTGGAGGCTGTGGCTCACTCAGAAAGAGAGAAAGAGGAGGAGGAGG	2938

OY	4019	ATGCCCTTCATTTTTCGTTACCATGATCTATTTATATGAGAGTGGGAGTGAAC	4078
DB	44179	ATGCCCTTCATTTTTCGTTACCATGATCTATTTATATGAGAGTGGGAGTGAAC	44238
OY	4079	CAAACTGCGGCAAGTTGGGACATCTACAGACCAAGTTATCTGCTAGAAATCCCC	4138
DB	44239	CAAACTGCGGCAAGTTGGGACATCTACAGACCAAGTTATCTGCTAGAAATCCCC	44298
OY	4139	TCTCATTGAGTTGGGAGAACTGCTGCTGGGGGCTTCAGGCTCTGTTGAGAGAG	4198
DB	44299	TCTCATTGAGTTGGGAGAACTGCTGCTGGGGGCTTCAGGCTCTGTTGAGAGAG	44358
OY	4199	GATCCTTTGATAGGGGATGACCTAGTCTATGTTGTTACTATCTCCTGCTAGTAA	4258
DB	44359	GATCCTTTGATAGGGGATGACCTAGTCTATGTTGTTACTATCTCCTGCTAGTAA	44418
OY	4259	AGCTGGAAGTAAACCCAGGAGCGCCAGAGTTCTTACAGTTGTACCCAGAACAA	4318
DB	44419	AGCTGGAAGTAAACCCAGGAGCGCCAGAGTTCTTACAGTTGTACCCAGAACAA	44478
OY	4319	CAAGACAGTATGATGCAAGATAGTGTAGTGTGGGAGAAAGAACTTAACCCCA	4378
DB	44479	CAAGACAGTATGATGCAAGATAGTGTAGTGTGGGAGAAAGAACTTAACCCCA	44538
OY	4379	AGGCCACAGTTCCGTTCCCTAGTTACATGCTGAGTGTGCTACTACTATGGG	4438
DB	44539	AGGCCACAGTTCCGTTCCCTAGTTACATGCTGAGTGTGCTACTACTATGGG	44598
OY	4439	TCTGAGTTGTCATGCTACAGATGATGATGTTCTATGCTGTATATATCTGAGC	4498
DB	44599	TCTGAGTTGTCATGCTACAGATGATGATGTTCTATGCTGTATATATCTGAGC	44658
OY	4499	ACTTGGAGGCTGAAGCAGAGATGCTATATGTTTGGGCGACGCTGAGCTATGAGC	4558
DB	44659	ACTTGGAGGCTGAAGCAGAGATGCTATATGTTTGGGCGACGCTGAGCTATGAGC	44718
OY	4559	GAGACTTCTCTTAAAGAAAAATGAAAAAGCCAGAGTGTGTGCGACAGCCCTTATCC	4618
DB	44719	GAGACTTCTCTTAAAGAAAAATGAAAAAGCCAGAGTGTGTGCGACAGCCCTTATCC	44778
OY	4619	CAGCACTTGGGAGGAGAACAGCAGAGATTTCTGAGTCTAAGGCCAGCCTGCTATAGA	4678
DB	44779	CAGCACTTGGGAGGAGAACAGCAGAGATTTCTGAGTCTAAGGCCAGCCTGCTATAGA	44838
OY	4679	GTCAGTTCCAGGACAGCCAGGCTCTACAGAGAAACCCCTGTTTGAAGAAACAGAAAC	4738
DB	44839	GTCAGTTCCAGGACAGCCAGGCTCTACAGAGAAACCCCTGTTTGAAGAAACAGAAAC	44898
OY	4739	AAAAAACAACCAACCAACCAACCAACCAACCAACCAACCTCTCATCTCTCATCTC	4798
DB	44899	AAAAAACAACCAACCAACCAACCAACCAACCAACCAACCTCTCATCTCTCATCTC	44958
OY	4799	TCTAGGCTGTCTGTCTGTAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT	4858
DB	44959	TCTAGGCTGTCTGTCTGTAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT	45018
OY	4859	CCTTTTATCATGCTGTCAGAGACGAGAAAGTTTCTGCTGGGACACATGGGACCTGA	4918
DB	45019	CCTTTTATCATGCTGTCAGAGACGAGAAAGTTTCTGCTGGGACACATGGGACCTGA	45078
OY	4919	GAAGTACTCTCTTGGCAGCCAAAAATTTCTGGGAAGGCTTCTGTGAAGAGTGTGCCG	4978
DB	45079	GAAGTACTCTCTTGGCAGCCAAAAATTTCTGGGAAGGCTTCTGTGTGTGTGTGTGCCG	45138
OY	4979	AT 4980	
DB	45139	AT 45140	

RESULT 3  
AC132885  
LOCUS AC132885 170074 bp DNA linear HFG 12-MAR-2003

DEFINITION Mus musculus clone RP24-263015, WORKING DRAFT SEQUENCE. 9 unordered pieces.  
AC132885  
VERSION AC132885.3  
KEYWORDS GI:28927760  
SOURCE HTGS\_PHASE1, HTGS\_DRAFT.  
ORGANISM Mus musculus (house mouse)  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 170074)  
AUTHORS Birren, B., Nusbaum, C. and Lander, E.  
TITLE Mus musculus, clone RP24-263015  
JOURNAL Unpublished  
2 (bases 1 to 170074)  
AUTHORS Birren, B., Nusbaum, C., Lander, E., All, A., Allen, N., Anderson, S., Baran, N., Bastien, V., Bloom, T., Boguslavsky, L., Boukhgalter, B., Camarata, J., Chang, J., Chazaro, B., Choepel, Y., Collymore, A., Cook, A., Cooke, P., DeRellano, K., Dewar, K., Diaz, J., Dodge, S., Faro, S., Ferrel, P., Fitzgerald, M., Gage, D., Galagan, J., Gardyna, S., Gord, S., Graham, L., Grand-Pierre, N., Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, G., Maclean, C., Macdonald, P., Major, J., Matthews, C., McCarthy, M., Meldrim, J., Menus, L., Mihova, T., Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Raymond, C., Retta, R., Rise, C., Rogov, P., Roman, J., Roy, A., Schauer, S., Schnupack, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Talamas, J., Testaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE Direct Submission  
JOURNAL Submitted (04-SEP-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
3 (bases 1 to 170074)  
AUTHORS Birren, B., Nusbaum, C., Lander, E., Abouelleil, A., Allen, N., Anderson, S., Arachchi, H. M., Baran, N., Bastien, V., Bloom, T., Boguslavsky, L., Boukhgalter, B., Camarata, J., Chang, J., Choepel, Y., Collymore, A., Cook, A., Cooke, P., Corbin, B., DeRellano, K., Diaz, J., Dodge, S., Dooley, K., Dorris, L., Erickson, J., Faro, S., Ferrel, P., Fitzgerald, M., Gage, D., Galagan, J., Gardyna, S., Graham, L., Grand-Pierre, N., Hafez, N., Hagopian, D., Hagos, B., Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, G., Lui, A., Mabbitt, R., Maclean, C., Macdonald, P., Major, J., Manning, J., Matthews, C., McCarthy, M., Meldrim, J., Menus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Rachpka, A., Ramasamy, U., Raymond, C., Retta, R., Rise, C., Rogov, P., Roman, J., Schauer, S., Schnupack, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Talamas, J., Testaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H., Vankataraman, V. S., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE Direct Submission  
JOURNAL Submitted (12-MAR-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
COMMENT All repeats were identified using RepeatMasker:  
http://ftp.genome.washington.edu/RM/RepeatMasker.html

Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR  
Web site: http://www-seq.wi.mit.edu  
Contact: sequence\_submissions@genome.wi.mit.edu  
Project Information  
Center project name: L27142  
Center clone name: 263.O.15  
Sequencing vector: Plasmid; n/a; 100% of reads



QY 3853 CATCTGAATGGTTTCTTTGTGTACACATTCCTCCCTGACACTCCTGGGAGGTCTGATCCTT 3912  
 1139 CATCTGAATGGTTTCTTTGTGTACACATTCCTCCCTGACACTCCTGGGAGGTCTGATCCTT 1198  
 QY 3913 GGCACATGATCTCGGATGTAACTGACGCCACCAAGAGAGAGAGGAGAGTCTGAGAGC 3972  
 1199 GGCACATGATCTCGGATGTAACTGACGCCACCAAGAGAGAGAGGAGAGTCTGAGAGC 1258  
 QY 3973 TGTGCTTGGGCTTTAGGCTGTGACATTCACCTCTTCCCTGAAATGGCCCTCCATT 4032  
 1259 TGTGCTTGGGCTTTAGGCTGTGACATTCACCTCTTCCCTGAAATGGCCCTCCATT 1318  
 QY 4033 TTTCCGTTACATGATCTATTTTATATCAAGAGTGGGAGTGAAGCCAACTGCCACA 4092  
 1319 TTTCCGTTACATGATCTATTTTATATCAAGAGTGGGAGTGAAGCCAACTGCCACA 1378  
 QY 4093 AGTTTGGACTACTGACACCAAGGTTATCTGCTCAAGAAATCCCTGTCTCTTGAAGTT 4152  
 1379 AGTTTGGACTACTGACACCAAGGTTATCTGCTCAAGAAATCCCTGTCTCTTGAAGTT 1438  
 QY 4153 GGGAGATCTGCTTGGGAGCTTCCAGGCTTGTAGAGGAGGATCTTGTATTA 4212  
 1439 GGGAGATCTGCTTGGGAGCTTCCAGGCTTGTAGAGGAGGATCTTGTATTA 1498  
 QY 4213 GGGCATGACCTAGTCTATGTTTACTATGATTCCTGTCAGTTAAAGCTGAACATAAA 4272  
 1499 GGGCATGACCTAGTCTATGTTTACTATGATTCCTGTCAGTTAAAGCTGAACATAAA 1558  
 QY 4273 CCCACGGACGCCCGACAGATTTCTTACAGTTGTTACCCCAAGCAACACAGACATGATTA 4332  
 1559 CCCACGGACGCCCGACAGATTTCTTACAGTTGTTACCCCAAGCAACACAGACATGATTA 1618  
 QY 4333 TGCAGAGATGAGTGTGGGAGAGAAAGAACTTAAACCCCAAGGCCACAGATTTC 4392  
 1619 TGCAGAGATGAGTGTGGGAGAGAAAGAACTTAAACCCCAAGGCCACAGATTTC 1678  
 QY 4393 CGTTCCCTAGTTCACATGCTCATGATGATGCTAGTACTATGGGCTGTGATGTTGATGC 4452  
 1679 CGTTCCCTAGTTCACATGCTCATGATGATGCTAGTACTATGGGCTGTGATGTTGATGC 1738  
 QY 4453 TACACATGATGATGTTTCATGTTGATGTTGATGTTGATGTTGATGTTGATGTTGATGTT 4512  
 1739 TACACATGATGATGTTTCATGTTGATGTTGATGTTGATGTTGATGTTGATGTTGATGTT 1798  
 QY 4513 AGCAGAGATGCTATGTTGATGTTGATGTTGATGTTGATGTTGATGTTGATGTTGATGTT 4572  
 1799 AGCAGAGATGCTATGTTGATGTTGATGTTGATGTTGATGTTGATGTTGATGTTGATGTT 1858  
 QY 4573 AAGAAAAAATGAAAGCCAGCAGTGTGGCAGACGCTTTAATCCAGCATTGGGAGG 4632  
 1859 AAGAAAAAATGAAAGCCAGCAGTGTGGCAGACGCTTTAATCCAGCATTGGGAGG 1918  
 QY 4633 CAGAACGAGGATTTCTGATGTTCAAGGCCAGCCTGTTATAGAGTGTGATGTTCAAGGAC 4692  
 1919 CAGAACGAGGATTTCTGATGTTCAAGGCCAGCCTGTTATAGAGTGTGATGTTCAAGGAC 1978  
 QY 4693 AGCCAGGCTTACACAGAGAAACCTGTTTGAAGAAACAGAAACCAAAACCAAAACCAAA 4752  
 1979 AGCCAGGCTTACACAGAGAAACCTGTTTGAAGAAACAGAAACCAAAACCAAAACCAAA 2038  
 QY 4753 CAAACCAAAACCAAAACCAAAACCAAAACCTCTCATCTCTCATCTCTCTGAGCTGTCT 4812  
 2039 CAAACCAAAACCAAAACCAAAACCAAAACCTCTCATCTCTCTCATCTCTCTGAGCTGTCT 2098  
 QY 4813 GTCTAGTGTGATGTTGGGAGCTTCAAGACTATATATTAAGGCTTTTATACATG 4872  
 2099 GTCTAGTGTGATGTTGGGAGCTTCAAGACTATATATTAAGGCTTTTATACATG 2158  
 QY 4873 GTCTAGTGTGATGTTCAAGACTTGTGAGACAGTGTGAGGAGCTTGAAGAAAGTACTCTT 4932  
 2159 GTCTAGTGTGATGTTCAAGACTTGTGAGACAGTGTGAGGAGCTTGAAGAAAGTACTCTT 2218

QY 4933 CCAGCCCCAAAATTTCTGGGAAGGCTTCTGAGGAGAGTGTCTCCGAT 4980  
 Db 2219 CCAGCCCCAAAATTTCTGGGAAGGCTTCTGAGGAGAGTGTCTCCGAT 2266

RESULT 4  
 BV067438/c 779 bp DNA linear STS 31-May-2003  
 LOCUS S212P60465FAL.T0 C2ECHII/El Mus musculus STS genomic, sequence  
 DEFINITION tagged site.

ACCESSION BV067438  
 VERSION BV067438.1 GI:31183233  
 KEYWORDS STS.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 779)  
 Made,C.M., Kulbokas,E.J. III, Kirby,A.W., Zody,M.C., Mullikin,J.C.,  
 Lander,E.S., Lindblad-Toh,K. and Daly,M.J.  
 The mosaic structure of variation in the laboratory mouse genome  
 Nature 420 (6915), 574-578 (2002)

TITLE JOURNAL  
 MEDLINE 22354684  
 PUBMED 12466852  
 COMMENT Contact: Kerstin Lindblad-Toh  
 Whitehead Institute for Biomedical Research, Center for Genome  
 Research  
 320 Charles Street, Cambridge, MA 02141, USA  
 Tel: 617/2521477  
 Fax: 617/2580903  
 Email: kersli@genome.wi.mit.edu  
 Primer A: None  
 Primer B: None  
 STS size: 779

Protocol:  
 WGS-discovery: Paired-end low-coverage whole genome shotgun reads  
 were generated from 12951/SVIMJ, C3H/HeJ, and BALB/cByJ. The WGS  
 reads were placed uniquely on the MGS3V3 C57BL/6J assembly and SNP  
 detection was carried out by SSAHA-SNP. 225,000 reads were  
 annotated  
 as STSs and 81,000 SNPs were annotated with alleles from C57BL/6J  
 and the strain from which the particular read came. The validation  
 rate for these SNPs was estimated at approximately 98%.

FEATURES  
 source location/Qualifiers  
 1..779  
 /organism="Mus musculus"  
 /mol\_type="genomic DNA"  
 /strain="C2ECHII/El"  
 /db\_xref="taxon:10090"  
 /map="1 31-720 75128449-75127762"  
 /clone\_lib="C2ECHII/El"  
 <1..>779

BASE COUNT 162 a 230 c 153 g 234 t  
 ORIGIN  
 Query Match 2.0%; Score 180; DB 11; Length 779;  
 Best Local Similarity 100.0%; Pred. No. 8,6e-87;  
 Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4597 TGTGTGCACAGCCTTTAATCCAGCAGCTTGGAGGACGAGACGAGATTTCTGAGTT 4656  
 Db 678 TGTGTGCACAGCCTTTAATCCAGCAGCTTGGAGGACGAGACGAGATTTCTGAGTT 619  
 QY 4657 CAAGGCGAGCTGTGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4716  
 Db 618 CAAGGCGAGCTGTGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 559  
 QY 4717 TGTGTGAAAAACACAGAAAAACCAAAACCAAAACCAAAACCAAAACCAAAACCAAAACCC 4776  
 Db 558 TGTGTGAAAAACACAGAAAAACCAAAACCAAAACCAAAACCAAAACCAAAACCAAAACCC 499

RESULT 5  
AC101018 149109 bp DNA linear HTG 21-AUG-2002  
LOCUS Rattus norvegicus clone RP32-328P7, \*\*\* SEQUENCING IN PROGRESS \*\*\*,  
DEFINITION 9 unordered pieces.  
AC101018  
AC101018 GI:22381320  
AC101018.2  
HTG: HTGS\_PHASE1; HTGS\_FULLTOP; HTGS\_ACTIVEFIN.  
KEYWORDS Rattus norvegicus (Norway rat)  
SOURCE Rattus norvegicus  
ORGANISM Mammalia; Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi;  
Eukaryota; Metazoa; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus  
1 (bases 1 to 149109)  
Birren,B., Nusbaum,C. and Lander,E.  
Rattus norvegicus, clone RP32-328P7  
Unpublished  
2 (bases 1 to 149109)  
Birren,B., Linton,L., Nusbaum,C., Lander,E., All,A., Allen,N.,  
Anderson,S., Barna,N., Bastien,V., Boguslavsky,L., Boukhgalter,B.,  
Brown,A., Camarata,J., Campolano,A., Chang,U., Chazaro,B.,  
Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,  
Cook,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,  
Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S.,  
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-pierre,N.,  
Hagos,B., Heatord,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,  
Jones,C., Kamat,A., Karatas,A., Kells,C., Larocque,K.,  
Lamacares,R., Landers,T., Lehoczy,J., Levine,R., Liu,G.,  
Maclean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,  
McCarthy,M., McEwan,P., McKernan,K., MCPheeters,R., Meldrum,J.,  
Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,  
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,  
Oliver,J., Peterson,K., Phunhhang,P., Pierre,N., Pollara,V.,  
Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,  
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupack,R.,  
Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,  
Strauss,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J.,  
Topham,K., Travers,M., Travis,N., Triggillo,J., Vassiliev,H.,  
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,  
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.  
Direct Submission  
Submitted 23-NOV-2001 Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
3 (bases 1 to 149109)  
Birren,B., Nusbaum,C., Lander,E., All,A., Allen,N., Anderson,S.,  
Barna,N., Bastien,V., Bloom,T., Boguslavsky,L., Boukhgalter,B.,  
Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,  
Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S.,  
Faro,S., Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J.,  
Gardyna,S., Gord,S., Graham,L., Grand-pierre,N., Hagos,B.,  
Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,  
Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,  
Liu,G., Maclean,C., Macdonald,P., Major,J., Matthews,C.,  
McCarthy,M., Meldrum,J., Meneus,L., Mihova,T., Mlenga,V.,  
Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,  
O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K.,  
Phunhhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,  
Roman,J., Roy,A., Schauer,S., Schupack,R., Seaman,S., Severy,P.,  
Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,  
Testaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,  
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,  
Zembek,L., Zimmer,A. and Zody,M.  
Direct Submission  
Submitted 21-AUG-2002 Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Aug 21, 2002 this sequence version replaced gi:17059792.  
All repeats were identified using RepeatMasker:  
http://ftp.genome.washington.edu/RM/RepeatMasker.html  
----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WITB  
Web site: http://www-seq.wi.mit.edu

Contact: sequence\_submissions@genome.wi.mit.edu  
----- Project Information  
Center project name: L14794  
Center clone name: 328\_P-7  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 9 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
1 46829: contig of 46829 bp in length  
\* 46830 46929: gap of 100 bp  
\* 46930 49621: contig of 2692 bp in length  
\* 49622 49721: gap of 100 bp  
\* 49722 70009: contig of 20288 bp in length  
\* 70010 70109: gap of 100 bp  
\* 70110 77196: contig of 7087 bp in length  
\* 77197 77296: gap of 100 bp  
\* 77297 99453: contig of 22157 bp in length  
\* 99454 99553: gap of 100 bp  
\* 99554 102879: contig of 3326 bp in length  
\* 102880 102979: gap of 100 bp  
\* 102980 120352: contig of 17373 bp in length  
\* 120353 120452: gap of 100 bp  
\* 120453 131546: contig of 11094 bp in length  
\* 131547 131646: gap of 100 bp  
\* 131647 149109: contig of 17463 bp in length.  
Location/Qualifiers  
1..149109  
/organism="Rattus norvegicus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10116"  
/clone="RP32-328P7"  
/clone\_11b="RPC1-32 Male Rat BAC"  
BASE COUNT 39036 a 34316 c 35185 g 39719 t 853 others  
ORIGIN  
Query Match 1.1%; Score 96; DB 2; Length 149109;  
Best Local Similarity 100.0%; Pred. No. 1.2e-40;  
Matches 96; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 5953 GAGTTTCAGAGGTAAAGCCATTTATTCACACATGCTCAAAATCAAG 6012  
Db 102361 GAGTTTACAGAGGTAAAGCCATTTATTCACACATGCTCAAAATCAAG 102420  
QY 6013 TGCAGAGAGATTAGCTGAGAGATGGGCTGTCA 6048  
Db 102421 TGCAGAGAGATTAGCTGAGAGATGGGCTGTCA 102456  
RESULT 6  
AC115173 193472 bp DNA linear HTG 15-NOV-2002  
LOCUS Rattus norvegicus clone CH230-286017, WORKING DRAFT SEQUENCE, 3  
DEFINITION 9 unordered pieces.  
AC115173  
AC115173 GI:25006753  
HTG: HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_FULLTOP.  
KEYWORDS Rattus norvegicus (Norway rat)  
SOURCE Rattus norvegicus  
ORGANISM Mammalia; Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi;  
Eukaryota; Metazoa; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus  
1 (bases 1 to 193472)  
Muzny,D., Marie., Metzker,M., Lee., Abramzon,S., Adams,C., Alder,J.,  
Allen,C., Allen,H., Alstbrooks,S., Amin,A., Anguiano,D.,  
Anyalebechi,V., Ayagi,A., Ayodeji,M., Baca,E., Baden,H.,  
Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,  
Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,  
Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,

Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J., Cleeland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L., Davila,M.L., Davis,C., Day-Carroll,L., De Anda,C., Dederich,D., Delgado,O., Denison,S., Deramo,C., Ding,Y., Dinh,H., Divya,K., Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K., Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G., Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P., Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M., Georgegeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W., Gunaratne,P., Haaland,W., Hamill,C., Hamilton,C., Hamilton,K., Harvey,Y., Haylak,P., Hawes,A., Henderson,N., Hernandez,D., Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hognes,M., Hollins,B., Howells,S., Huliyk,S., Hume,J., Idelblid,D., Jackson,A., Jackson,L., Jacob,L., Jiang,N., Johnson,B., Johnson,R., Jolivet,A., Karathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C., Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J., Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J., Lorensuhewa,L., Louisedge,H., Lozado,R.J., Lu,X., Ma,J., Maheshwari,M., Mahindaratne,M., Mahmoud,M., Malloy,K., Mangum,A., Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E., Manthey,S., Mcleod,M.P., McNeill,T.Z., Meenen,E., Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S., Morgan,M., Morris,K., Morris,S., Mundaya,M., Murphy,M., Nair,L., Nantervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Nwockelameh,O., Okwono,G., Olarnpunsagoon,A., Pal,S., Parks,K., Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkuch,C., Plopper,F., Polidexter,A., Popovic,D., Plums,E., Pu,L.-L., Puzo,M., Qutroz,J., Rachlin,E., Reeves,K., Reigler,M.A., Reigh,R., Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F., Rivers,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J., Sanders,M., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H., Shetty,J., Shvartsbeyn,A., Sisson,I., Sletter,C.D., Smajs,D., Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J., Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabori,P., Taylor,C., Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usman,K., Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J., Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F., Williams,G., Willson,R., Wlarczyk,R., Wooden,H., Worley,K., Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V., Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O., Weinstock,G. and Gibbs,R.A.

Unpublished  
2 (bases 1 to 193472)  
Direct Submission  
Worley,K.C.  
Submitted (15-MAR-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 193472)  
Rat Genome Sequencing Consortium.  
Direct Submission  
Submitted (15-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
On Nov 15, 2002 this sequence version replaced gi:23678661.  
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: hgsc-help@bcm.tmc.edu  
----- Project Information  
Center project name: CH230-286017  
----- Summary Statistics  
Assembly program: Phrap; version 0.990329  
Consensus quality: 185211 bases at least Q40  
Consensus quality: 186794 bases at least Q30  
Consensus quality: 188105 bases at least Q20  
Estimated insert size: 184863; sum-of-contigs estimation  
Quality coverage: 8x in Q20 bases; sum-of-contigs estimation  
-----  
\* NOTE: Estimated insert size may differ from sequence length  
\* (see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 3 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
\* 1 190920: contig of 190920 bp in length  
\* 190921 191020: gap of unknown length  
\* 191021 192270: contig of 1250 bp in length  
\* 192271 192370: gap of unknown length  
\* 192371 193472: contig of 1102 bp in length.  
Location/Qualifiers  
1..193472  
/organism="Rattus norvegicus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10116"  
/clone="CH230-286017"  
1..1118  
/note="wgs\_end\_extension  
clone\_end:T7"  
4580..5464  
/note="clone\_boundary  
clone\_end:T7  
site:  
end\_sequence:RXAHM937VC"  
4668..5345  
/note="clone\_boundary  
clone\_end:T7  
site:  
end\_sequence:RXAHM937VB"  
187253..188705  
/note="wgs\_end\_extension  
clone\_end:T7"  
189469..190920  
/note="wgs\_end\_extension  
clone\_end:T7"  
BASE COUNT 51460 a 44772 c 44401 g 48651 t 4188 others  
ORIGIN  
Query Match 1.1%; Score 96; DB 2: Length 193472;  
Best local similarity 100.0%; Pred. No. 1.2e-40;  
Matches 96; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 5953 GAGTTTACAGAGGTAAGGCGCAGCATTTATTCACAGACATGACTCAAAATCAAG 6012  
|||||  
DB 125203 GAGTTTACAGAGGTAAGGCGCAGCATTTATTCACAGACATGACTCAAAATCAAG 125262  
|||||  
QY 6013 TTCAAGGAGAGATTAGCTTGAGAGATGGGGCTGTCA 6048  
|||||  
DB 125263 TTCAAGGAGAGATTAGCTTGAGAGATGGGGCTGTCA 125298  
|||||



RESULT 7  
AC136730/c 145489 bp DNA linear HTG 22-FEB-2003  
LOCUS AC136730.3  
DEFINITION Mus musculus clone RP23-76E22, WORKING DRAFT SEQUENCE, 4 unordered  
pieces  
ACCESSION AC136730 GI:28467347  
VERSION AC136730.3  
KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 145489)  
REFERENCE 1 Birren, B., Nusbaum, C. and Lander, E.  
TITLE Mus musculus, clone RP23-76E22  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 145489)  
AUTHORS Birren, B., Nusbaum, C., Lander, E., All, A., Allen, N., Anderson, S.,  
Barna, N., Bastien, V., Bloom, T., Boguslavsky, L., Boukhalter, B.,  
Camata, J., Chang, J., Chazaro, B., Choepel, Y., Collymore, A.,  
Cook, A., Cooke, P., Deatellano, K., Dewar, K., Diaz, J. S., Dodge, S.,  
Faro, S., Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J.,  
Gardyna, S., Gord, S., Graham, L., Grand-Pierre, N., Hafez, N.,  
Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,  
Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R.,  
Lindblad-Toh, K., Liu, G., Liu, A., Mabbitt, R., Maclean, C.,  
Macdonald, P., Major, J., Manning, J., Matthews, C., McCarthy, M.,  
Meldrum, J., Meneus, L., Mihova, T., Menega, V., Murphy, T., Naylor, J.,  
Nguyen, C., Nicol, R., Norbu, C., O'Connor, T., O'Donnell, P.,  
O'Neill, D., Oliver, J., Peterson, K., Phunkhag, P., Plante, R.,  
Rachpaka, A., Ramasamy, U., Raymond, C., Retta, R., Risse, C.,  
Roman, J., Schauer, S., Schnupack, R., Seaman, S., Severy, P., Smith, C.,  
Spencer, B., Stange-Thomann, N., Stojanovic, N., Stubbs, M.,  
Talamas, J., Testaye, S., Theodore, J., Topham, K., Travers, M.,  
Vassiliev, H., Venkataraman, V. S., Viel, R., Vo, A., Wilson, B., Wu, X.,  
Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.  
TITLE Direct Submission  
JOURNAL Submitted (07-NOV-2002) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
REFERENCE 3 (bases 1 to 145489)  
AUTHORS Birren, B., Nusbaum, C., Lander, E., Abouelell, A., Allen, N.,  
Anderson, S., Arachchi, H. M., Barna, N., Bastien, V., Bloom, T.,  
Boguslavsky, L., Boukhalter, B., Camata, J., Chang, J., Choepel, Y.,  
Collymore, A., Cook, A., Cooke, P., Corum, B., Deatellano, K.,  
Diaz, J. S., Dodge, S., Dooley, K., Dorris, L., Erickson, J., Faro, S.,  
Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J., Gardyna, S.,  
Graham, L., Grand-Pierre, N., Hafez, N., Hagopian, D., Hagos, B.,  
Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,  
Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R.,  
Lindblad-Toh, K., Liu, G., Liu, A., Mabbitt, R., Maclean, C.,  
Macdonald, P., Major, J., Manning, J., Matthews, C., McCarthy, M.,  
Meldrum, J., Meneus, L., Mihova, T., Menega, V., Murphy, T., Naylor, J.,  
Nguyen, C., Nicol, R., Norbu, C., O'Connor, T., O'Donnell, P.,  
O'Neill, D., Oliver, J., Peterson, K., Phunkhag, P., Plante, R.,  
Rachpaka, A., Ramasamy, U., Raymond, C., Retta, R., Risse, C.,  
Roman, J., Schauer, S., Schnupack, R., Seaman, S., Severy, P., Smith, C.,  
Spencer, B., Stange-Thomann, N., Stojanovic, N., Stubbs, M.,  
Talamas, J., Testaye, S., Theodore, J., Topham, K., Travers, M.,  
Vassiliev, H., Venkataraman, V. S., Viel, R., Vo, A., Wilson, B., Wu, X.,  
Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.  
TITLE Direct Submission  
JOURNAL Submitted (22-FEB-2003) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
COMMENT On Feb 22, 2003 this sequence version replaced gi:28412006.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RM/RepeatMasker.html  
----- Genome Center  
Center: Whitehead Institute/MIT Center for Genome Research  
Center code: WtBR  
Web site: http://www-seq.wi.mit.edu  
Contact: sequence\_submissions@genome.wi.mit.edu  
----- Project Information

Center project name: L26307  
Center clone name: 76\_E-22  
----- Summary Statistics  
Sequencing vector: Plasmid: n/a; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.960731  
Consensus quality: 144239 bases at least Q40  
Consensus quality: 144835 bases at least Q30  
Consensus quality: 145007 bases at least Q20  
Insert size: 145000; agarose-fp  
Insert size: 145189; sum-of-ctgls  
Quality coverage: 8.6 in Q20 bases; agarose-fp  
Quality coverage: 8.6 in Q20 bases; sum-of-ctgls  
-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 4 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
\* 1 14246: contig of 14246 bp in length  
\* 14247 14346: gap of 100 bp  
\* 14347 16056: contig of 1710 bp in length  
\* 16057 16156: gap of 100 bp  
\* 16157 11985: contig of 95829 bp in length  
\* 11986 112085: gap of 100 bp  
\* 112086 145489: contig of 33404 bp in length.  
\* location/Qualifiers  
1. 145489  
/organism="Mus musculus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10090"  
/clone="RP23-76E22"  
/clone\_lib="PCF-23 Female Mouse BAC"  
1. 14246  
/note="assembly\_fragment  
clone\_end:SP6  
vector\_side:left"  
14347. 16056  
/note="assembly\_fragment"  
16157. 11985  
/note="assembly\_fragment"  
112086. 145489  
/note="assembly\_fragment  
clone\_end:T7  
vector\_side:right"  
BASE COUNT 40514 a 34404 c 33449 g 36801 t 321 others  
ORIGIN  
Query Match 0.98; Score 82; DB 2; Length 145489;  
Best Local Similarity 100.0%; Pred. No. 5.9e-33;  
Matches 82; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 4638 GCAGGAGATTTCGATTGAGCCAGCTGCTTACAGTGTCCAGACGCA 4697  
AC115850  
LOCUS AC115850  
DEFINITION Mus musculus clone RP24-252010, WORKING DRAFT SEQUENCE, 4 unordered  
pieces  
ACCESSION AC115850  
VERSION AC115850.4 GI:28867059  
KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
SOURCE Mus musculus (house mouse)







```

misc_feature      60990..71936
                  /note="assembly-fragment"
misc_feature      72037..88642
                  /note="assembly-fragment"
misc_feature      88743..101771
                  /note="assembly-fragment"
misc_feature      101872..111509
                  /note="assembly-fragment"
misc_feature      115610..138341
                  /note="assembly-fragment"
misc_feature      138442..159683
                  /note="assembly-fragment"
misc_feature      159784..173892
                  /note="assembly-fragment"
                  clone_end:T7
                  vector_side:right"
BASE COUNT      41036 a 45314 c 42508 g 43019 t 2015 others
ORIGIN
Query Match      0.9%; Score 82; DB 2; Length 173892;
Best Local Similarity 100.0%; Pred. No. 5.9e-33;
Matches 82; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      4593 GCAGTGTGGTGCACACGCTTTAATCCACGACTTGGAGGACAGACGACGAGATTCTCG 4652
         |||||||
Db      146185 GCAGTGTGGTGCACACGCTTTAATCCACGACTTGGAGGACAGACGACGAGATTCTCG 146126
         |||||||

QY      4653 AGTTCAAGCGCCGCTGTCTA 4674
         |||||||
Db      146125 AGTTCAAGCGCCGCTGTCTA 146104
         |||||||

RESULT 10
AC139041/c      189241 bp      DNA      linear      HTG 17-MAR-2003
LOCUS
DEFINITION
SEQUENCE: 6 unordered pieces.
AC139041
AC139041.3 GI:28976013
HTG: HTGS_PHASE1: HTGS_DRAFT.
SOURCE
ORGANISM
Mus musculus (house mouse)
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 189241)
Birren,B., Nusbaum,C. and Lander,E.
Mus musculus chromosome 14, clone RP23-283B12
Unpublished
2 (bases 1 to 189241)
Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
Boguslavsky,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y.,
Collymore,A., Cook,A., Cooke,P., Corum,B., Dearellano,K.,
Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,
Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J., Gardyna,S.,
Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
Lindblad-Toh,K., Liu,G., Lui,A., Mabbitt,R., Maclean,C.,
Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
Melidim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J.,
Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
O'Neill,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
Rachpka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
Roman,J., Schauer,S., Schuback,R., Seaman,S., Severy,P., Smith,C.,
Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
Talamas,J., Testaye,S., Theodore,J., Topham,K., Travers,M.,
Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (23-JAN-2003) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 189241)

```

```

AUTHORS
Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
Boguslavsky,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y.,
Collymore,A., Cook,A., Cooke,P., Corum,B., Dearellano,K.,
Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,
Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J., Gardyna,S.,
Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
Lindblad-Toh,K., Liu,G., Lui,A., Mabbitt,R., Maclean,C.,
Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
Melidim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J.,
Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
O'Neill,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
Rachpka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
Roman,J., Schauer,S., Schuback,R., Seaman,S., Severy,P., Smith,C.,
Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
Talamas,J., Testaye,S., Theodore,J., Topham,K., Travers,M.,
Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (17-MAR-2003) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 17, 2003 this sequence version replaced gi:28603968.
All repeats were identified using RepeatMasker:
http://ftp.genome.washington.edu/RM/RepeatMasker.html

COMMENT
JOURNAL
TITLE
JOURNAL
COMMENT
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: 283_B-12
Center clone name: 283_B-12
----- Summary Statistics
Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 187878 bases at least Q40
Consensus quality: 188076 bases at least Q40
Consensus quality: 188202 bases at least Q20
Insert size: 175000; agarose-fp
Quality coverage: 13.1 in Q20 bases; agarose-fp
Quality coverage: 12.1 in Q20 bases; sum-of-coverage
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 6 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 48250: contig of 48250 bp in length
* 48251 48350: gap of 100 bp
* 48351 53618: contig of 5268 bp in length
* 53619 53718: gap of 100 bp
* 53719 62174: contig of 8456 bp in length
* 62175 62274: gap of 100 bp
* 62275 109247: contig of 46973 bp in length
* 109248 109347: gap of 100 bp
* 109348 148402: contig of 39055 bp in length
* 148403 148502: gap of 100 bp
* 148503 189241: contig of 40739 bp in length.
*
* Location/Qualifiers
* 1..189241
* /organism="Mus musculus"
* /mol_type="genomic DNA"
* /db_xref="taxon:10090"
* /chromosome="14"
* /map="14"
FEATURES
source

```

```

misc_feature      /clone="Rp23-283b12"
                  .48250
                  /note="assembly_fragment
clone_end:SP6
vector_side:left"
misc_feature      48351..53618
                  /note="assembly_fragment"
misc_feature      53719..62174
                  /note="assembly_fragment"
misc_feature      62275..109247
                  /note="assembly_fragment"
misc_feature      109348..148402
                  /note="assembly_fragment"
misc_feature      148503..189241
                  /note="assembly_fragment
clone_end:T7
vector_side:right"
BASE COUNT      48741 a 45049 c 45485 g 49461 t 505 others
ORIGIN
Query Match      0.9%; Score 82; DB 2; Length 189241;
Best Local Similarity 100.0%; Pred. No. 5.9e-33;
Matches 82; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 4638 GCAGGACATTTCGAGTTCAGGCGCCTGCTCTATAGAGTCCAGACGCCA 4697
Db 143086 GCAGGACATTTCGAGTTCAGGCGCCTGCTCTATAGAGTCCAGACGCCA 143027
OY 4698 GGCGTACACAGAGAAACCTCT 4719
Db 143026 GGCGTACACAGAGAAACCTCT 143005

RESULT 11
AC131335/c
LOCUS             AC131335             190776 bp    DNA             linear    HTG 08-MAR-2003
DEFINITION       Mus musculus clone Rp23-32D9, WORKING DRAFT SEQUENCE, 9 unordered
                  pieces.
ACCESSION        AC131335
VERSION          AC131335.4 GI:28882340
KEYWORDS         HTG: HTGS PHASE1; HTGS: DRAFT.
SOURCE           Mus musculus (house mouse)
ORGANISM         Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 190776)
Birren,B., Nusbaum,C. and Lander,E.
Mus musculus, clone Rp23-32D9
Unpublished
2 (bases 1 to 190776)
Birren,B., Nusbaum,C., Lander,E., All,A., Allen,N., Anderson,S.,
Barn,N., Bastien,V., Bloom,T., Boguslavsky,L., Boukhalter,B.,
Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collamore,A.,
Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J.,
Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,
Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kanat,A.,
Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,
Liu,G., Maclean,C., Macdonald,P., Major,J., Matthews,C.,
McCarthy,M., Mceldrim,J., Meneus,L., Mihova,T., Mienga,V.,
Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K.,
Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
Roman,J., Roy,A., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,
Tefaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
Zembek,L., Zimmer,A. and Zody,M.
TITLE
JOURNAL
REFERENCE
3 (bases 1 to 190776)

```

```

AUTHORS
Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
Anderson,S., Arachchi,H.M., Barn,N., Bastien,V., Bloom,T.,
Boguslavsky,L., Boukhalter,B., Camarata,J., Chang,J., Choepel,Y.,
Collamore,A., Cook,A., Cooke,K., Doris,B., Dearellano,K.,
Diaz,J.S., Dodge,S., Dooley,K., Faros,J., Erickson,J., Faro,S.,
Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J., Gardyna,S.,
Graham,L., Grand-Pierre,N., Hafe,N., Hagopian,D., Hagos,B.,
Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
Kanat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
Lindblad-Toh,K., Liu,G., Lui,A., Mabbitt,R., Maclean,C.,
Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
Mceldrim,J., Meneus,L., Mihova,T., Mienga,V., Murphy,T., Naylor,J.,
Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
O'Neill,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
Talamas,J., Tefaye,S., Theodore,J., Topham,K., Travers,M.,
Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
TITLE
JOURNAL
COMMENT
Direct Submission
Submitted (08-MAR-2003) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 8, 2003 this sequence version replaced gi:25956385.
All repeats were identified using RepeatMasker:
Smit,A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L24187
Center clone name: 32_D_9
----- Summary Statistics
Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 188840 bases at least 40
Consensus quality: 189359 bases at least 30
Consensus quality: 189634 bases at least 20
Insert size: 18200; agarose-ftp
Insert size: 189976; sum-of-contigs
Quality coverage: 8.4 in Q20 bases; agarose-ftp
Quality coverage: 8.1 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 9 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 34490..contig of 34490 bp in length
* 34491 34590: gap of 100 bp
* 34591 36121: contig of 3531 bp in length
* 36122 38222: gap of 100 bp
* 38222 43662: contig of 5441 bp in length
* 43662 43762: gap of 100 bp
* 43762 45771: gap of 5809 bp in length
* 45771 49671: gap of 100 bp
* 49671 58834: contig of 9163 bp in length
* 58834 58935: gap of 100 bp
* 58935 77918: contig of 18984 bp in length
* 77918 78018: gap of 100 bp
* 78018 127564: contig of 43546 bp in length
* 127564 127665: gap of 100 bp
* 127665 167916: contig of 40252 bp in length
* 167917 168016: gap of 100 bp
* 168017 190776: contig of 22760 bp in length.
Location/Qualifiers

```

```
source
1. .190776
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/clone="RP23-32D9"
/clone_11b="RP23-32D9"
1. .34490
/feature="assembly-fragment"
clone_end:sp6
vector_side:left"
misc_feature
34591..38121
/feature="assembly-fragment"
misc_feature
38222..43662
/feature="assembly-fragment"
misc_feature
43763..49571
/feature="assembly-fragment"
misc_feature
49672..58834
/feature="assembly-fragment"
misc_feature
58935..77918
/feature="assembly-fragment"
misc_feature
78019..127564
/feature="assembly-fragment"
misc_feature
127665..167916
/feature="assembly-fragment"
misc_feature
168017..190776
/feature="assembly-fragment"
clone_end:t7
vector_side:right"
BASE COUNT 57965 a 34254 c 34054 g 63703 t 800 others
ORIGIN
Query Match 0.9%; Score 82; DB 2; Length 190776;
Best Local Similarity 100.0%; Pred. No. 5.9e-33;
Matches 82; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 4638 GCAGGCAATTTCTGATTCAGGCCAGCCTGCTATAGAGTTCAGGACGCA 4697
DB 60883 GCAGGCAATTTCTGATTCAGGCCAGCCTGCTATAGAGTTCAGGACGCA 60824
QY 4698 GGGCTACAGAGAAACCTGT 4719
DB 60823 GGGCTACAGAGAAACCTGT 60802
RESULT 12
AC122900 203855 bp DNA linear ROD 21-SEP-2002
DEFINITION Mus musculus chromosome 15 clone RP23-39N24, complete sequence.
ACCESSION AC122900
VERSION AC122900.3 GI:23266394
KEYWORDS HTG.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 203855)
McPherson,J.D. and Waterston,R.H.
The sequence of Mus musculus clone
Unpublished
2 (bases 1 to 203855)
McPherson,J.D. and Waterston,R.H.
Direct Submission
Submitted (28-MAY-2002) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
3 (bases 1 to 203855)
McPherson,J.D. and Waterston,R.H.
Direct Submission
Submitted (01-SEP-2002) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
4 (bases 1 to 203855)
McPherson,J.D. and Waterston,R.H.
Direct Submission
Submitted (21-SEP-2002) Genome Sequencing Center, 4444 Forest Park
```

```
COMMENT
Parkway, St. Louis, MO 63108, USA
On Sep 21, 2002 this sequence version replaced gi:22476032.
----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
Contact: submissions@wustl.edu
----- Project Information -----
Center project name: M_BA0039N24
-----
FEATURES
source
1. .203855
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="15"
/clone="RP23-39N24"
BASE COUNT 61572 a 38691 c 38009 g 65583 t
ORIGIN
Query Match 0.9%; Score 82; DB 10; Length 203855;
Best Local Similarity 100.0%; Pred. No. 5.9e-33;
Matches 82; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 4593 GCAGTGTGGACACGCCCTTATCCAGCACTTGGAGGACAGACGACGATTTCTG 4652
DB 170522 GCAGTGTGGACACGCCCTTATCCAGCACTTGGAGGACAGACGACGATTTCTG 170581
QY 4653 AGTTCAAGGCCAGCCTGTGCTTA 4674
DB 170582 AGTTCAAGGCCAGCCTGTGCTTA 170603
RESULT 13
AC127697 222185 bp DNA linear HTG 22-JUL-2002
LOCUS AC127697
DEFINITION Mus musculus chromosome UNK clone RP24-83110, WORKING DRAFT
SEQUENCE. 13 unordered pieces.
AC127697.2 GI:21913472
VERSION AC127697.2
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 222185)
McPherson,J.D. and Waterston,R.H.
The sequence of Mus musculus clone
Unpublished
2 (bases 1 to 222185)
McPherson,J.D. and Waterston,R.H.
Direct Submission
Submitted (18-JUL-2002) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
3 (bases 1 to 222185)
McPherson,J.D. and Waterston,R.H.
Direct Submission
Submitted (22-JUL-2002) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
On Jul 21, 2002 this sequence version replaced gi:21903681.
----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
Contact: submissions@wustl.edu
----- Project Information -----
Center project name: M_BB0083110
-----
----- Summary Statistics -----
Sequencing vector: M13; 0%
Sequencing vector: plasmid; 100%
```

Chemistry: Dye-primer ET; 0% of reads  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Assembly program: Phrap; version 0.990319  
 Consensus quality: 217367 bases at least Q40  
 Consensus quality: 218557 bases at least Q30  
 Consensus quality: 219259 bases at least Q20

\* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 13 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

```

1 1042: contig of 1042 bp in length
* 1043 1142: gap of unknown length
* 1143 3141: contig of 1999 bp in length
* 3142 3241: gap of unknown length
* 3242 7211: contig of 3970 bp in length
* 7212 7311: gap of unknown length
* 7312 10408: contig of 3097 bp in length
* 10409 10508: gap of unknown length
* 10509 17015: contig of 6507 bp in length
* 17016 17115: gap of unknown length
* 17116 27558: contig of 10443 bp in length
* 27559 27658: gap of unknown length
* 27659 41472: contig of 13814 bp in length
* 41473 41572: gap of unknown length
* 41573 58617: contig of 17045 bp in length
* 58618 58718: gap of unknown length
* 58718 83267: contig of 24549 bp in length
* 83267 83367: gap of unknown length
* 83367 130000: contig of 46634 bp in length
* 130001 172736: contig of 42636 bp in length
* 172737 172836: gap of unknown length
* 172837 221107: contig of 48271 bp in length
* 221108 221207: gap of unknown length
* 221208 222185: contig of 978 bp in length.
  
```

FEATURES

```

source
1..222185
  /organism="Mus musculus"
  /mol_type="genomic DNA"
  /db_xref="taxon:10090"
  /chromosome="UNK"
  /clone="RP24-83110"
misc_feature
1..1042
  /note="assembly_name:Contig25"
misc_feature
1143..3141
  /note="assembly_name:Contig26"
misc_feature
3242..7211
  /note="assembly_name:Contig27"
misc_feature
7312..10408
  /note="assembly_name:Contig28"
misc_feature
10509..17015
  /note="assembly_name:Contig29"
misc_feature
17116..27558
  /note="assembly_name:Contig30"
misc_feature
27659..41472
  /note="assembly_name:Contig31"
misc_feature
41573..58617
  /note="assembly_name:Contig32"
misc_feature
58718..83266
  /note="assembly_name:Contig33"
misc_feature
83367..130000
  /note="assembly_name:Contig34"
misc_feature
130101..172736
  /note="assembly_name:Contig35"
misc_feature
172837..221107
  /note="assembly_name:Contig36"
misc_feature
221208..222185
  /note="assembly_name:Contig37"
  
```

BASE COUNT 67728 a 41550 c 41568 g 70139 t 1200 others

Query Match 0.9% Score 82: DB 2: Length 222185;  
 Best Local Similarity 100.0% Pred. No. 5.9e-33;  
 Matches 82: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

OY 4593 GCAGTGTGACACAGCCCTTAATCCAGCAGCTGGAGGACGACGAGATTCTG 4652
    ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 26396 GCAGTGTGACACAGCCCTTAATCCAGCAGCTGGAGGACGACGAGATTCTG 26455
OY 4653 AGTTCAGGCCAGCCTGTCTTA 4674
    ||||||||||||||||||||||||
Db 26456 AGTTCAGGCCAGCCTGTCTTA 26477
  
```

```

RESULT 14
AC079490/c 226909 bp DNA linear HTG 02-SEP-2000
LOCUS Mus musculus clone RP23-190A2, WORKING DRAFT SEQUENCE, 53 unordered
DEFINITION pieces.
ACCESSION AC079490
VERSION AC079490.1 GI:9964855
KEYWORDS HTG, HTGS, PHASE1, HTGS, DRAFT.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 (bases 1 to 226909)
DOE Joint Genome Institute.
TITLE Sequencing of Mouse
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 226909)
REFERENCE
DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (02-SEP-2000) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov
  
```

Project Information  
 Center Project Name: 1809834  
 Center clone name: RP23-190A2  
 -----  
 Summary Statistics  
 Consensus quality: 180161 bases at least Q40  
 Consensus quality: 198601 bases at least Q30  
 Consensus quality: 203733 bases at least Q20  
 Estimated insert size: 209000; agarose-1p estimation  
 Estimated insert size: 221709; sum-of-contigs estimation  
 Quality coverage: 5.84 in Q20 bases; sum-of-contigs estimation  
 Quality coverage: 5.84 in Q20 bases; sum-of-contigs estimation.  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 53 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

```

1 1120: contig of 1120 bp in length
* 1121 1220: gap of unknown length
* 1221 2574: contig of 1354 bp in length
* 2575 2674: gap of unknown length
* 2675 4062: contig of 1388 bp in length
* 4063 4162: gap of unknown length
* 4163 5666: contig of 1504 bp in length
* 5667 5767: gap of unknown length
* 5767 6896: contig of 1129 bp in length
* 6896 8025: gap of unknown length
* 8025 8025: contig of 1030 bp in length
  
```

8026 8125: gap of unknown length  
 8126 9388: contig of 1263 bp in length  
 9389 9488: gap of unknown length  
 9489 10710: contig of 1222 bp in length  
 10711 10810: gap of unknown length  
 10811 11811: contig of 1001 bp in length  
 11812 11911: gap of unknown length  
 11912 12965: contig of 1054 bp in length  
 12966 13065: gap of unknown length  
 13066 14268: contig of 1203 bp in length  
 14269 14368: gap of unknown length  
 14369 15846: contig of 1478 bp in length  
 15847 15946: gap of unknown length  
 15947 17219: contig of 1273 bp in length  
 17220 17319: gap of unknown length  
 17320 18672: contig of 1353 bp in length  
 18673 18772: gap of unknown length  
 18773 19999: contig of 1227 bp in length  
 20000 20099: gap of unknown length  
 20100 21459: contig of 1360 bp in length  
 21460 21559: gap of unknown length  
 21560 22622: contig of 1063 bp in length  
 22623 22722: gap of unknown length  
 22723 23726: contig of 1004 bp in length  
 23727 23826: gap of unknown length  
 23827 25149: contig of 1323 bp in length  
 25150 25249: gap of unknown length  
 25250 26839: contig of 1590 bp in length  
 26840 26939: gap of unknown length  
 26940 28630: contig of 1691 bp in length  
 28631 28730: gap of unknown length  
 28731 30265: contig of 1535 bp in length  
 30266 31703: contig of 1338 bp in length  
 31704 31803: gap of unknown length  
 31804 32951: contig of 1148 bp in length  
 32952 33051: gap of unknown length  
 33052 35329: contig of 2278 bp in length  
 35330 35429: gap of unknown length  
 35430 38496: contig of 3067 bp in length  
 38497 38596: gap of unknown length  
 38597 40521: contig of 1925 bp in length  
 40522 40621: gap of unknown length  
 40623 42212: contig of 1591 bp in length  
 42213 44441: gap of unknown length  
 44442 44541: gap of unknown length  
 44542 46397: contig of 1856 bp in length  
 46398 46497: gap of unknown length  
 46498 48685: contig of 2188 bp in length  
 48686 48785: gap of unknown length  
 50811 50911: contig of 2026 bp in length  
 50912 52640: gap of unknown length  
 52641 52740: gap of unknown length  
 52741 53952: contig of 1212 bp in length  
 53953 54052: gap of unknown length  
 54053 58018: contig of 3966 bp in length  
 58019 58118: gap of unknown length  
 58119 61436: contig of 3318 bp in length  
 61437 61536: gap of unknown length  
 61537 64032: contig of 2496 bp in length  
 64033 64132: gap of unknown length  
 64133 66752: contig of 2620 bp in length  
 66753 66852: gap of unknown length  
 66853 70097: contig of 3245 bp in length  
 70098 70197: gap of unknown length  
 70199 73503: contig of 3306 bp in length  
 73504 73603: gap of unknown length  
 73604 80346: contig of 6743 bp in length  
 80347 80446: gap of unknown length  
 80447 86466: contig of 6020 bp in length  
 86467 86566: gap of unknown length

86567 92278: contig of 5712 bp in length  
 92279 92378: gap of unknown length  
 92379 98374: contig of 6496 bp in length  
 98375 98875: gap of unknown length  
 98875 106050: contig of 7076 bp in length  
 106051 106150: gap of unknown length  
 106150 113625: contig of 7475 bp in length  
 113626 113725: gap of unknown length  
 113726 129075: contig of 15350 bp in length  
 129076 129175: gap of unknown length  
 129176 141800: contig of 12625 bp in length  
 141801 141900: gap of unknown length  
 141901 155581: contig of 13681 bp in length  
 155582 155681: gap of unknown length  
 155682 170324: contig of 14643 bp in length  
 170325 170424: gap of unknown length  
 170425 184681: contig of 14257 bp in length  
 184682 184781: gap of unknown length  
 184782 206159: contig of 21378 bp in length  
 206160 226909: gap of unknown length  
 226909 226909: contig of 20650 bp in length.

FEATURES  
 source  
 1..226909  
 /organism="Mus musculus"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:10090"  
 /clone="RP23-190A2"  
 /clone\_1b="RPCI mouse BAC library 23"

BASE COUNT 52802 a 57780 c 58798 g 52284 t 5245 others

ORIGIN

Query Match 0.9% Score 82; DB 2; Length 226909;  
 Best Local Similarity 100.0%; Pred. No. 5; 9e-33;  
 Matches 82; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4593 GCAGTGTGGACACACCCCTTTAATCCAGCACTTGGAGGACAGACAGCAGATTCTCG 4652  
 |||||||  
 DB 35815 GCAGTGTGGACACACCCCTTTAATCCAGCACTTGGAGGACAGACAGCAGATTCTCG 35756

QY 4653 AGTTCAAGCGCAGCTGGTCTA 4674  
 |||||||  
 DB 35755 AGTTCAAGCGCAGCTGGTCTA 35734

RESULT 15  
 AC141471/c  
 LOCUS 232568 bp DNA linear HTG 18-MAR-2003  
 DEFINITION Mus musculus chromosome UNK clone RP24-326K11, WORKING DRAFT  
 AC141471  
 AC141471.2 GI:29029419  
 VERSION  
 ACCESSION HTG; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_FULLTOP.  
 KEYWORDS  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 232568)  
 McPherson, J.D. and Waterston, R.H.  
 The sequence of Mus musculus clone  
 unpublished  
 2 (bases 1 to 232568)  
 McPherson, J.D. and Waterston, R.H.  
 Direct Submission  
 Submitted (16-MAR-2003) Genome Sequencing Center, 4444 Forest Park  
 Parkway, St. Louis, MO 63108, USA  
 3 (bases 1 to 232568)  
 McPherson, J.D. and Waterston, R.H.  
 Direct Submission  
 Submitted (18-MAR-2003) Genome Sequencing Center, 4444 Forest Park  
 Parkway, St. Louis, MO 63108, USA  
 On Mar 18, 2003 this sequence version replaced gi:28975171.  
 ----- Genome Center -----

```
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
Contact: submissions@wustl.wustl.edu
-----
Project Information
Center project name: M_BB0326K11
-----
----- Summary Statistics -----
Sequencing vector: M13: %
Sequencing vector: Plasmid: %
Chemistry: Dye-terminator ET: % of reads
Chemistry: Dye-terminator Big Dye: % of reads
Assembly program: Phrap: version 0.990319
Consensus quality: 22068 bases at least Q40
Consensus quality: 22401 bases at least Q30
Consensus quality: 22351 bases at least Q20
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 20 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 1364: contig of 1364 bp in length
* 1365 1464: gap of unknown length
* 1465 3212: contig of 1748 bp in length
* 3213 3312: gap of unknown length
* 3313 5171: contig of 1859 bp in length
* 5172 5272: gap of unknown length
* 5272 7005: contig of 1734 bp in length
* 7006 7106: gap of unknown length
* 7106 9867: contig of 2762 bp in length
* 9868 13925: contig of 3958 bp in length
* 13926 14025: gap of unknown length
* 14026 17200: contig of 3175 bp in length
* 17201 17300: gap of unknown length
* 17301 24239: contig of 6939 bp in length
* 24240 24339: gap of unknown length
* 24340 31078: contig of 6739 bp in length
* 31079 31179: gap of unknown length
* 31179 38186: contig of 7008 bp in length
* 38187 38286: gap of unknown length
* 38287 47404: contig of 9118 bp in length
* 47405 47504: gap of unknown length
* 47505 59212: contig of 11708 bp in length
* 59213 59312: gap of unknown length
* 59313 71856: contig of 12544 bp in length
* 71857 71956: gap of unknown length
* 71957 89788: contig of 17832 bp in length
* 89789 89888: gap of unknown length
* 89889 107217: contig of 17329 bp in length
* 107218 107317: gap of unknown length
* 107318 121884: contig of 14567 bp in length
* 121885 121984: gap of unknown length
* 121985 140588: contig of 18604 bp in length
* 140589 140688: gap of unknown length
* 140689 163356: contig of 22668 bp in length
* 163357 163456: gap of unknown length
* 163457 192246: contig of 28790 bp in length
* 192247 192346: gap of unknown length
* 192347 233568: contig of 40222 bp in length.
*
Location/Qualifiers
1. .232568
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="UNK"
/clone="RP24-326K11"
1. .1364
/misc_feature
/note="assembly_name:Contig17"
```

```
misc_feature 1465. .3212
/note="assembly_name:Contig19"
misc_feature 3313. .5171
/note="assembly_name:Contig21"
misc_feature 5272. .7005
/note="assembly_name:Contig22"
misc_feature 7106. .9867
/note="assembly_name:Contig25"
misc_feature 9968. .13925
/note="assembly_name:Contig26"
misc_feature 14026. .17200
/note="assembly_name:Contig27"
misc_feature 17301. .24239
/note="assembly_name:Contig28"
misc_feature 24340. .31078
/note="assembly_name:Contig29"
misc_feature 31179. .38186
/note="assembly_name:Contig30"
misc_feature 38287. .47404
/note="assembly_name:Contig31"
misc_feature 47505. .59212
/note="assembly_name:Contig32"
misc_feature 59313. .71856
/note="assembly_name:Contig33"
misc_feature 71957. .89788
/note="assembly_name:Contig34"
misc_feature 89889. .107217
/note="assembly_name:Contig35"
misc_feature 107318. .121884
/note="assembly_name:Contig36"
misc_feature 121985. .140588
/note="assembly_name:Contig37"
misc_feature 140689. .163356
/note="assembly_name:Contig38"
misc_feature 163457. .192246
/note="assembly_name:Contig39"
misc_feature 192347. .232568
/note="assembly_name:Contig40"
BASE COUNT 54848 a 58265 c 57016 g 60272 t 2167 others
ORIGIN
Query Match 0.9% Score 82: DB 2: Length 232568;
Best Local Similarity 100.0% Pred. No. 5.9e-33;
Matches 82: Conservative 0: Mismatches 0: Indels 0: Gaps 0:
QY 4638 GCAGCAGATTTCGATTCAGGCCAGCCTGCTCTATAGAGTCAAGTCCAGACACCA 4697
|||||
Db 118996 GCAGCAGATTTCGATTCAGGCCAGCCTGCTCTATAGAGTCAAGTCCAGACACCA 118937
QY 4698 GGGCTACACAGAGAACCCTGT 4719
|||||
Db 118936 GGGCTACACAGAGAACCCTGT 118915
```

Search completed: September 13, 2003, 21:35:34  
Job time : 31324 secs